

OTHER INFORMATION: n=a, g, c or t

Seq ID no. 129

Query Match 6.5%; Score 35.8; DB 4; Length 3001;
Best Local Similarity 54.1%; Pred. No. 0.9;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 419 TGAATGCAACATTTTCCATTAATCTGAACCTGACGATGACCTGAAAAATTTAGCTGA 478
DB 2385 TGCAGATATCATTTTCCAGCTGCTGAAAGCTAGAGATCAAGAAAAACAACAA 2326
QY 479 AAAAACTGATCTATGCGCAAGCATTTTGTATCATCTTTACTTTCCTGCTGCAAG 538
DB 2325 AAACTCGTGAATTTAGCAGCTGACTATAGGCTTGATCAATTTGAACTTCCCTGACATT 2266
QY 539 GAGAACTGATGAG 553
DB 2265 GAGTATGATTTGAG 2251

RESULT 4

US-08-545-528D-1
Sequence 1, Application US/08545528D
Patent No. 6537773
GENERAL INFORMATION:
APPLICANT: Frazer et al.
TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
PATENT NO. 6537773
TITLE OF INVENTION: Thee of, and Uses Thee of
FILE REFERENCE: PB193PI
CURRENT APPLICATION NUMBER: US/08/545,528D
PRIORITY FILING DATE: 1995-10-19
PRIORITY APPLICATION NUMBER: US 08/488,018
PRIORITY FILING DATE: 1995-06-07
PRIORITY APPLICATION NUMBER: US 08/473,545
PRIORITY FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 580073
TYPE: DNA
ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 6.2%; Score 34.4; DB 4; Length 580073;
Best Local Similarity 48.9%; Pred. No. 17;
Matches 92; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 366 TAAAGTCACTCATATGTTATACCAACGCTTTCTGTCGCACTATCAGCATGATGC 425
DB 281067 TGAATTAACATGATTAACCAACCAATTTTAAACCACTTGATGCTTAAAGC 281126
QY 426 AACATTTCTTAATATCTGAACCTGACGATGCTGAAAAATTTAGCTGAAAAAAC 485
DB 281127 AACATTTCTGATATGATGATGATGATGATGATGATGATGATGATGATGATG 281186
QY 486 CTGATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545
DB 281187 AAACGTAATTAACATTTCTTTTAAACGATTTGAAAGATGATGATGATGATGATG 281246
QY 546 GAATGAG 553
DB 281247 AAAAGAG 281254

RESULT 5

US-09-134-001C-2533
Sequence 2533, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIORITY FILING DATE: 1997-11-08
PRIORITY APPLICATION NUMBER: US 60/064,964
PRIORITY FILING DATE: 1997-11-08
PRIORITY APPLICATION NUMBER: US 60/055,779
PRIORITY FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2533
LENGTH: 1500
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2533

Query Match 6.1%; Score 33.8; DB 4; Length 1500;
Best Local Similarity 54.4%; Pred. No. 2.6;
Matches 68; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 102 CAGATACGAATACCTCTCTTCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 161
DB 146 CAGGCGCTATGCTTCTTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 205
QY 162 CAGCACTTTTCCGCAAGGCTGAAAAATGCTTCAAGTCTGCTGCGCATCTGCTT 221
DB 206 CTGTAAGTATGAGGCTGCTGAAAAATTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 265
QY 222 GTCTT 226
DB 266 GTT 270

RESULT 6

US-09-280-116-144
Sequence 144, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035000/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
PRIORITY FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 144
LENGTH: 246
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: trypsin-like serine proteases
US-09-280-116-144

Query Match 6.1%; Score 33.6; DB 4; Length 246;
Best Local Similarity 52.9%; Pred. No. 1.5;
Matches 72; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 286 CCAATGAGAACTCTGAGACATCTATATGACCCCAATTAATCTCTCTCTGATG 345
DB 38 CCAATGAGAGACAGAGGCTCTCCCAATGACCCAGGATATGATGATGATGATGATGATG 97
QY 346 ATCCATTTCCCTTGAATTTAAGTCACTCATATGATATACCAACAGCTTTCTGTT 405
DB 98 CTGCTCCAGAGTCAATTTGCTGTTATCTATCTATCTATCTATCTATCTATCTATCTAT 157
QY 406 CCACTATCAACATGA 421
DB 158 CCGGTTGCAATCATGA 173

RESULT 7

US-09-497-855A-37
Sequence 37, Application US/09497855A
Patent No. 6605432
GENERAL INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2004, 20:57:37 ; Search time 2538 Seconds

(without alignments)
9443.931 Million cell updates/sec

Title: US-09-531-113-5981

Perfect score: 553
Sequence: 1 gtacgtatcttcacagacatt.....ccaagagagactgatgag 553

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: 1: gb_ba: 2: gb_bg: 3: gb_in: 4: gb_cm: 5: gb_ov: 6: gb_pat: 7: gb_pl: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_ro: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_mu: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_ro: 28: em_ro: 29: em_ro: 30: em_vl: 31: em_vl: 32: em_vl: 33: em_vl: 34: em_vl: 35: em_vl: 36: em_vl: 37: em_vl: 38: em_vl: 39: em_vl: 40: em_vl: 41: em_vl:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	133.8	24.2	1004	8 AK066056	AK066056 Oryza sat
C 2	125.8	22.7	1088	8 AK065722	AK065722 Oryza sat
C 3	123.2	22.3	591	8 AY098968	AY098968 Arabidops
C 4	123.2	22.3	789	8 AF462824	AF462824 Arabidops
C 5	98.6	17.8	277	6 AR245689	AR245689 Sequence
C 6	58	10.5	105858	2 AP004324	AP004324 Oryza sat
C 7	56.4	10.2	41087	8 AB024029	AB024029 Arabidops
C 8	53.6	9.7	105967	8 AP005647	AP005647 Oryza sat
C 9	53.6	9.7	133387	2 AP004061	AP004061 Oryza sat
C 10	53.2	9.6	7218	6 I66494	I66494 Sequence 14
C 11	45.2	8.2	163277	2 AC074259	AC074259 Trypanoso
C 12	45.2	8.2	252394	3 AE014833	AE014833 Plasmodiu
C 13	45	8.1	126038	2 AC016068	AC016068 Homo sapi
C 14	43.2	7.8	564	11 BV089971	BV089971 RPAMSE00
C 15	42	7.6	218313	2 AC112682	AC112682 Mus muscu
C 16	42	7.6	307150	1 CNSPAX01	AU248283 Pyrococcu
C 17	42	7.6	349980	6 AX041106	AX041106 Sequence
C 18	41.8	7.6	408	6 AX300942	AX300942 Sequence
C 19	41.6	7.5	192541	2 AC079165	AC079165 Mus muscu
C 20	41.6	7.5	198181	2 AC035248	AC035248 Mus muscu
C 21	41.6	7.5	215063	2 AC107763	AC107763 Mus muscu
C 22	41.2	7.5	256880	2 AC095430	AC095430 Rattus no
C 23	41	7.4	267434	2 AC095185	AC095185 Rattus no
C 24	41	7.4	267434	2 AC128998	AC128998 Rattus no
C 25	40.8	7.4	224956	2 AC019008	AC019008 Mus muscu
C 26	40.8	7.4	231316	2 AC115138	AC115138 Rattus no
C 27	40.8	7.4	251096	10 AL586255	AL586255 Mouse DNA
C 28	40.2	7.3	1141	6 AX083744	AX083744 Sequence
C 29	40.2	7.3	16876	2 BX255937	BX255937 Danio rer
C 30	40.2	7.3	181925	2 AC020710	AC020710 Homo sapi
C 31	40.2	7.3	197864	9 AC021956	AC021956 Homo sapi
C 32	40	7.2	53385	2 AC136994	AC136994 Homo sapi
C 33	40	7.2	125733	10 AC117250	AC117250 Mus muscu
C 34	40	7.2	140176	9 AC022188	AC022188 Homo sapi
C 35	40	7.2	152779	9 AC018943	AC018943 Homo sapi
C 36	40	7.2	187842	9 AC114778	AC114778 Homo sapi
C 37	40	7.2	217619	2 AC135511	AC135511 Mus muscu
C 38	39.8	7.2	164603	2 AC119466	AC119466 Rattus no
C 39	39.8	7.2	223996	2 AC094504	AC094504 Rattus no
C 40	39.6	7.2	142565	10 AL772205	AL772205 Mouse DNA
C 41	39.6	7.2	188713	2 AC128580	AC128580 Rattus no
C 42	39.6	7.2	243075	2 AC097827	AC097827 Rattus no
C 43	39.4	7.1	42942	7 AB045978	AB045978 Staphyloc
C 44	39.4	7.1	43095	6 BD245282	BD245282 Developme
C 45	39.4	7.1	111873	9 AC073275	AC073275 Homo sapi

ALIGNMENTS

RESULT 1
AK066056/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AK066056 1004 bp mRNA
Oryza sativa (japonica cultivar-group) cDNA clone:J013052F01, full
insert sequence.

AK066056
AK066056.1 GI:32976074
FLI_CDNA; CAP trapper.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

The Rice Full-length cDNA Consortium, National Institute of

TITLE

Agrobiological Sciences Rice Full-Length cDNA Project Team:
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yasaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Ohnuki, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group: Otsu, Y., Murakami, K.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusunagi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN:
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Komano, H., Miyazaki, A., Otsu, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice

JOURNAL

Science 301 (5631), 376-379 (2003)

MEDLINE

22752273

PUBMED

12869764

AUTHORS

2 (bases 1 to 1004)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hara, A., Hashidume, W.,
Hayashizaki, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T.,
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kikuchi, S., Kishikawa, Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Komano, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusunagi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Nishikura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohta, Y., Ohtsuki, K., Oka, M., Ooka, H.,
Otsu, N., Ota, Y., Otsu, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, T., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yasaki, J., Yokomizo, S., and
Yoshimura, A.

TITLE

Direct Submission

JOURNAL

Submitted (05-DEC-2001)

PUBMED

Shohei Kikuchi, National Institute of

AUTHORS

Agrobiological Sciences, Department of Molecular Genetics, Head of

PUBMED

Laboratory of Gene Expression: 2-1-2 Kannondai, Tsukuba, Ibaraki

PUBMED

305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,

PUBMED

Tel:81-29-858-7007, Fax:81-29-858-7007)

PUBMED

This clone is one of the 28K full-length cDNA clones from japonica

PUBMED

rice.

PUBMED

URL: http://cdna01.dna.affrc.go.jp/cDNA/

PUBMED

NIAS Rice Full-length cDNA Project Team:

PUBMED

Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yasaki, J.,

PUBMED

Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,

PUBMED

Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohnuki, K., Shishiki, T., and

PUBMED

Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otsu, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kodama, T., Kurosaki, T., Kusunagi, T., Lu, M., Masuda, H., Miura, J.,
Mizuno, K., Narikawa, R., Nishikura, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Sugiyama, A., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K., and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hara, A.,
Hara, A., Hashizume, W., Hayashizaki, K., Hayatsu, N., Hiramoto, K.,
Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Komano, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Miura, J.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohta, Y.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,

FEATURES

SOURCE

Location/Qualifiers
1. 1004
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/DB_xref="taxon:39947"
/clone="J013052P01"

ORIGIN

Query Match

Best Local Similarity

Matches

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

Query

Db

JOURNAL

MEDLINE

PUBMED

Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takahara-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A., and Hayashizaki, Y.
Location/Qualifiers
1. 1004
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/DB_xref="taxon:39947"
/clone="J013052P01"

REFERENCE

AUTHORS

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

PUBMED

JOURNAL

MEDLINE

PUBMED

PUBMED
REFERENCE
AUTHORS

12869764
2 (bases 1 to 1088)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imanura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Kono,H., Kouda,M., Koyama,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Maeda,H., Matsubara,K., Matsuyama,T., Mura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Oka,H., Oseko,N., Ota,Y., Otsomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,M., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 28K full-length cDNA clones from japonica rice.
URL: http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otsomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Maeda,H., Mura,J., Mizuno,K., Narikawa,R., Nishikura,J., Oka,M., Sano,H., Sugano,S., Yoshimura,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hori,F., Hotta,I., Iida,J., Imanura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Kono,H., Kouda,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Oseko,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Saeki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,T., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.
Location/Qualifiers
1. 1088
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013038J22"

FEATURES
source

ORIGIN

Query Match 22.7%; Score 125.8; DB 8; Length 1088;
Best Local Similarity 64.0%; Pct. No. 11e-19;
Matches 190; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

OY 133 TATCATCTGAGTTCCTGTCGAATGCAACACTTTTCCGGAACGCTGTAAAT 192
DB 837 TACCATCTTGAGTCTCTTCGAAATGCGCAGTTTCTCCGATAGGTGAAGGAG 778

OY 193 AGCTCAAGTCGTTGGCGATCTGTGTTGTTTGATGCAAGATACCTTGCTCACTC 252
DB 777 GACCTGAGAGCTCTGCAATATCGCTCTTGTCGTGATGAGGAGCCCTTGTCATCGGTA 718
OY 253 AGCTCAATTCACCTGGAAGTCAATCTAGTTCACCAATGGAATCTGGGACATCTTA 312
DB 717 AACCTACTCTTCACTCCAAAGCTCCAGCTCCAGCAAGAAAGAACTCGAATGTCACGA 658
OY 313 TGACCCCAATTAACCTCTCTCTCTGATGATGCATTCCTCCCTTGACTTTAAGTC 372
DB 657 TGCCCTTGACCTTCATCTTTCTTTCTTTAATTAACATTCACCTTAATTTCAAGCTT 598
OY 373 AACTCATATGATTAACAACAAGCTTCTGTTCCGAATATCACCATGAATCAACA 429
DB 597 AGTTCATGATATACCTACTCTCTTTATTAACAAGGTGACTGAAGAAGCATCA 541

RESULT 3
AY098968/c
LOCUS
DEFINITION
Arabidopsis thaliana AT5G58110/K2119_90 mRNA, complete cds.
ACCESSION
AY098968
VERSION
AY098968.1 GI:20466112
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosid; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 591)
Kim,C.J., Chen,H., Cheuk,R., Meyers,M.C., Shim,P., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamliya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.
Unpublished
2 (bases 1 to 591)
Kim,C.J., Chen,H., Cheuk,R., Meyers,M.C., Shim,P., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamliya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (22-APR-2002) Salk Institute Genomic Analysis Laboratory (SIGAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPD cDNAs (RAPD cDNA: 'RIKEN Arabidopsis Full-length cDNA': Seki,M., Narusaka,M., Ishida,J., Hayashizaki,Y. and Shinozaki,K.).

REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 591)
Kim,C.J., Chen,H., Cheuk,R., Meyers,M.C., Shim,P., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamliya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.
Unpublished
2 (bases 1 to 591)
Kim,C.J., Chen,H., Cheuk,R., Meyers,M.C., Shim,P., Banh,J., Bowser,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jones,T., Kamliya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (22-APR-2002) Salk Institute Genomic Analysis Laboratory (SIGAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPD cDNAs (RAPD cDNA: 'RIKEN Arabidopsis Full-length cDNA': Seki,M., Narusaka,M., Ishida,J., Hayashizaki,Y. and Shinozaki,K.).

COMMENT

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAPD cDNAs: Kim,C.J., Chen,H., Cheuk,R., Meyers,M.C., Shim,P., Banh,J., Bowser,L., Chang,E., Dale,J.M., Goldsmith,A.D., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Davis,R.W., Theologis,A., and Ecker,J.R.

Kim,C.J. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PI.

FEATURES

source

Location/Qualifiers
 1..591
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="U21471"
 /ecotype="Columbia"
 /note="This clone is in pUNI 51"

CDS

1..591
 /note="unknown protein"
 /codon_start=1
 /product="AT5g58110/K21119_90"
 /protein_id="AAU1978.1"
 /db_xref="GI:20466113"
 /translation="MERERETKRGSSRYVWRATSDAAPPIPKLSNNDVSLNTPASIGSLMNRAGTVEKSLTKWATDRLEKLGVSIGFSGKRIIDVNRQVADATLVTYNNKRVGVTEYELSLKVEGWSFEENMKVKSLTEIPFSFGEIDLDLVDVLSDDKELSQLKORIKLDMKQFLBPIRLKGLFPEGLKDR"

ORIGIN

Query Match 22.3%; Score 123.2; DB 8; Length 591;
 Best Local Similarity 63.5%; Pred. No. 5e-19;
 Matches 188; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 133 TATCTATCTGAGTCTCTGTTGCAATTCAGCAACTTTCCGAGACAGCTGTAAAT 192
 |||||
 Db 590 TATCTATCTTGAATCTCTGTTGCAAGCGCGAGTTGAGCGGATAGCTCCAGAAC 531
 QY 193 AGCTTCAAGTGTGGCATCTGTGCTCTGTTGATGCAAGATATCCCTTGCTTCACTC 252
 |||||
 Db 530 TGCTTCATATCCAACTTAATCCGCTGCTTCAAGCTGCGGAAAGCTCTTGTCTGCTG 471
 QY 253 AGTCTCATCTCCAGCTGCAATGCAATCTAGTCAACCAATGAGACTCGGACATCTATA 312
 |||||
 Db 470 AGCTTCATCATCTCCAGCTGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 411
 QY 313 TGACCCCAATAAATCTTCTCTCTCTGATGATTCATCTCCCTTGACTTTTAAGTTC 372
 |||||
 Db 410 CTCCTTCACCTTCTTCTATATCTCTTCAATGACCACTCTCCCTCAACTTTAAGGAA 351
 QY 373 AACTCATATGATTAACCAACAGCTTTCTTTCGAACTATCAACATTAATGCAAC 428
 |||||
 Db 350 AGCTCGTAAGTATAGCCAACTCTTCTTCTGTCGCAACTGCACTAAGAAAGCATC 295

RESULT 4

AF462824/c 789 bp mRNA linear PLN 08-JAN-2002

LOCUS AF462824
 DEFINITION Arabidopsis thaliana AT5g58110/K21119_90 mRNA, complete cds.
 ACCESSION AF462824
 VERSION AF462824.1 GI:18087564
 KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 789)
 Cheuk, R., Chen, H., Kim, C.-J., Meyers, M.C., Banh, J., Bowser, L.,

AUTHORS

Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,

Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,

Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,

Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,

Southwick, A., Tang, C.C., Tortum, M., Wu, H.C., Yamada, K.,

Yamamura, Y., Yu, G., Yu, S., Shinzaki, K., Davis, R.W., Theologis, A.,

and Ecker, J.R.
 Arabidopsis cDNA clones
 Unpublished
 2 (bases 1 to 789)
 Cheuk, R., Chen, H., Kim, C.-J., Meyers, M.C., Banh, J., Bowser, L.,

Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,

Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,

TITLE

JOURNAL

COMMENT

Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,
 Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,
 Southwick, A., Tang, C.C., Tortum, M., Wu, H.C., Yamada, K.,
 Yamamura, Y., Yu, G., Yu, S., Shinzaki, K., Davis, R.W., Theologis, A.,
 and Ecker, J.R.
 Direct Submission
 Submitted (27-DEC-2001) Salk Institute Genomic Analysis Laboratory
 (SIGAL), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10301 N. Torrey Pines Road, La Jolla, CA 92037,
 USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RFLP CDNA (RFLP CDNA : 'RIKEN
 Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y., and Shinzaki, K.

The Salk, Stanford, PESC (SSP) Consortium members carried out the
 sequencing and annotation of the RFLP CDNA: Cheuk, R., Chen, H.,
 Kim, C.J., Meyers, M.C., Shin, P., Banh, J., Bowser, L., Dale, J.M.,
 Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G., Lam, B.,
 Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M., Onodera, C.S.,
 Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A., Tang, C.C.,
 Tortum, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W.,
 Theologis, A., and Ecker, J.R.
 Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinzaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
 contributed equally to this work as PI.

FEATURES

source

Location/Qualifiers
 1..789
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="RAF109-09-K03 (R21471)"
 /ecotype="Columbia"
 1..33
 34..624
 /note="unknown protein"
 /codon_start=1
 /product="AT5g58110/K21119_90"
 /protein_id="AAU58914.1"
 /db_xref="GI:18087565"
 /translation="MERERETKRGSSRYVWRATSDAAPPIPKLSNNDVSLNTPASIGSLMNRAGTVEKSLTKWATDRLEKLGVSIGFSGKRIIDVNRQVADATLVTYNNKRVGVTEYELSLKVEGWSFEENMKVKSLTEIPFSFGEIDLDLVDVLSDDKELSQLKORIKLDMKQFLBPIRLKGLFPEGLKDR"

CDS

QY 133 TATCTATCTGAGTCTCTGTTGCAATTCAGCAACTTTCCGAGACAGCTGTAAAT 192
 |||||
 Db 623 TATCTATCTTGAATCTCTGTTGCAAGCGCGAGTTGAGCGGTATAGGCTCCAGAAC 564
 QY 193 AGCTTCAAGTGTGGCATCTGTGCTCTGTTGATGCAAGATATCCCTTGCTTCACTC 252
 |||||
 Db 563 TGCTTCATATCCAACTTAATCCGCTGCTTCAAGCTGCGGAAAGCTCTTGTCTTGGTG 504
 QY 253 AGTCTCATCTCCAGCTGCAATGCAATCTAGTCAACCAATGAGAACTGTGGACATCTATA 312
 |||||
 Db 503 AGCTTCATCATCTCCAGCTGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 444
 QY 313 TGACCCCAATAAATCTTCTCTCTCTGATGATTCATCTCCCTTGACTTTTAAGTTC 372
 |||||
 Db 443 CTCCTTCACCTTCTTCTATATCTCTTCAATGACCACTCTCCCTCAACTTTAAGGAA 384
 QY 373 AACTCATATGATTAACCAACAGCTTTCTTTCGAACTATCAACATGAATGCAAC 428
 |||||
 Db 383 AGCTCGTAAGTATAGCCAACTCTTCTTCTGTCGCAACTGCACTAAGAAAGCATC 328

ORIGIN

3'UTR
 Query Match 22.3%; Score 123.2; DB 8; Length 789;
 Best Local Similarity 63.5%; Pred. No. 4.9e-19;
 Matches 188; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

RESULT 5
AR245689/C 277 bp DNA linear PAT 20-DEC-2002

LOCUS AR245689 Sequence 1048 from patent US 6476212.
DEFINITION AR245689
ACCESSION AR245689
VERSION AR245689.1 GI:27293563
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL

Unknown.
Unclassified.
1 (bases 1 to 277)
Iaigudi,R.V., Ito,L.Y. and Sherman,B.K.
Polynucleotides and polypeptides derived from corn ear
Patent: US 6476212-A 1048 05-NOV-2002;
location/Qualifiers
1..277
/organism="unknown"
/mol_type="genomic DNA"

Query Match 17.8% Score 98.6; DB 6; Length 277;
Best Local Similarity 61.5% Pred.No.4e-13; Indels 0; Gaps 0;
Matches 158; Conservative 0; Mismatches 99;

Db 167 ACTTTCCCGAAGAGCTGTAAAAATAGCTTCAAGTCGTGGCAGTCGTGTCTTCTT 226
257 ATTTTCTCCTGGATGGGCGCAAGAGACTTGAGATCTCTGGAGATCCGTGTATCAT 198

Qy 227 GATGAGAGATATCCCTTCTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 286
197 CAGAGCGAGAGCCCTTGTATGTGCGGAGCTTACTTGTATCTTCAAGTCATGATGTGCG 138

Qy 287 CAATGAGAGACTGTGGGACATCTATATGACCCCAATTAAGCTTCTTCTCTCTGTATGA 346
137 CGATGAGAAATCTCGGAGATGTCTAGATGCCCTTGTATCTTCTGTGTTCTTCTTGATTA 78

Qy 347 TCCATTCCTCCCTTGTGACTTTAAGTCAACTCATATGATATGATTAACCAAGAGCTTCTTGTCC 406
77 ACCATTCACCCGTAAATCTCAAGCTTGTGTATGATTAATCTCACTCTTCTGTGTGC 18

Db 407 GAGCATGACCATGAAT 423
17 GGACATGACCTGAAT 1

Qy 407 GAGCATGACCATGAAT 423
17 GGACATGACCTGAAT 1

RESULT 6
AP004324/c 105858 bp DNA linear HTG 21-MAR-2002

LOCUS AP004324/c
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 6 clone
OJ1215_E11, *** SEQUENCING IN PROGRESS ***.
ACCESSION AP004324
VERSION AP004324.1 GI:16930109
KEYWORDS HTG, HTGS, PHASE2.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
REFERENCE Submitted (14-NOV-2001) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program, Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)

1
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare (GMS) genomic DNA, chromosome 6, BAC
clone:OJ1215_E11
Published Only in Database (2001)
2 (bases 1 to 105858)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (14-NOV-2001) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program, Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT

The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.

NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and the accession number will be preserved.

Location/Qualifiers

1. 105858

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="6"

/clone="OJ1215_E11"

ORIGIN

Query Match 10.5%; Score 58; DB 2; Length 105858;

Best Local Similarity 62.3%; Pred. No. 0.0013; Indels 0; Gaps 0;

Matches 91; Conservative 0; Mismatches 55;

Oy 132 CTATCTATCTCTAGTTCCTGTTCTGATTCAGCACTTTTCCGACAGCGCTGTAAAAA 191

Db 68554 CTATCTATCTCTAGTTCCTGTTCTGATTCAGCACTTTTCCGACAGCGCTGTAAAAA 68495

Oy 192 TAGCTTCAGTCGTTGGCGCATCTGTGCTCTTGTCTTGATCGAAGATATCCCTGCTTCACT 251

Db 68494 CGACTTCGAATCCCTTGCTGATTTGTGCTCTTGTGTGCGACGAGAGTCCCTGATGTGCGT 68435

Oy 252 CAGTCTCAGCTTCACCGTCGAAGCAT 277

Db 68434 AAAACTTATTTGCACCTGAAAGCAT 68409

RESULT 7

AB024029/c 41087 bp DNA linear PLN 27-DEC-2000

LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K21L19.

ACCESSION AB024029 BA000015

VERSION AB024029.1 GI:4519188

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (stee)

Sato, S., Nakamura, Y., Kanehiko, T., Katoh, T., Asamizu, E., Kotani, H. and Tabata, S.

REFERENCE

AUTHORS

TITLE

Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty BAC and TAC clones

JOURNAL DNA Res. 7 (1), 31-63 (2000)

MEDLINE 200125

PUBMED 10718197

REFERENCE 2 (bases 1 to 41087)

AUTHORS Nakamura, Y.

TITLE Direct Submission

JOURNAL Submitted (24-FEB-1999) Yasuhiro Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yata, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)

COMMENT

Addresses for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=21L19

Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.

LGGMONVNPFGGKMEQONCTMLNPNFENIEYPLPADNMVPRDNNRSKJLDESIM
 NPINDSOBYGATMTMDPEKMSGKPENDNDVDVDDIMNEMKKEBNNGVPAVATRF
 GFDSPFP
 complement (join(18421, .18675,18945, .18992))
 /note="gene_id:K21L19.8
 unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB11000.1"
 /db_xref="GI:9759534"
 /translation="MNSNKSRIIRMEPCDQIRDSTNKRIRGTCDSLHLIYGNMSR
 NRRKLIIMRIKGOINRYTILSFGLPVLPHLSTIGLMSSSPPLVLRKQIRF"
 complement (join(19771, .20875,21080, .21408))
 /note="gene_id:K21L19.9"
 /codon_start=1
 /evidence=not_experimental
 /product="glucanase; glucan endo-1,3-beta-glucosidase"
 /protein_id="BAB11001.1"
 /db_xref="GI:9759535"
 /translation="MNGSVLILLLAVALLCQASSIGANWGTQASHPLPDIIVAMLR
 ENGLOKVLPAEYDITLALGSGIEVWVGIPNEMLATIASLKAEKRVAKNVSITH
 STDVNIIRYVAVGNPEPLSTVNGSYLSTFPALRNIGAIKAGIQNGVKTCPINAD

CDS

Query Match 10.2%; Score 56.4; DB 8; Length 41087;
 Best Local Similarity 63.0%; Pred. No. 0.0034;
 Matches 87; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 133 TATCATCTGAGTTCCTGTTGCAATGACGAACTTTCCCGAAGAGCTGTAAAT 192
 Db 30693 TATCATCTTGAAGTCTGTTGCAAGAGCGGAGCTTTGAGCGGATAGCTCCCAAGAC 30694

QY 193 AGCTTCAAGTCGTTGGCAGATCTGTCTTGTCTTGATGCAAGATATCCCTGCTTCACTC 252
 Db 30633 TGCTTCAATCCAACTTAATCGCTGCTTCAAGCTGCGAAGCTCTTGTCTTGGCTG 30574

QY 253 AGTCTCACTTCACCTGCG 270
 Db 30573 AGCTTCACTTCACCTGCG 30556

RESULT 8
 AP005647
 LOCUS
 DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
 BAC clone:OSUNBA002E05, complete sequence.
 ACCESSION AP005647
 VERSION AP005647.2 GI:31414614
 KEYWORDS HTG;
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
 AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.
 TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
 clone:OSUNBA002E05
 JOURNAL Published Only in Database (2002)
 REFERENCE 2 (bases 1 to 105967)
 AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.
 JOURNAL Direct Submission
 Submitted (21-AUG-2002) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rsgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 COMMENT On Jun 4, 2003 this sequence version replaced gi:22415832.
 The orientation of the sequence is from -21M13 to M13rev of the BAC
 clone.

FEATURES
 source
 1..105967
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"

ORIGIN
 Query Match 9.7%; Score 53.6; DB 8; Length 105967;
 Best Local Similarity 61.4%; Pred. No. 0.015;
 Matches 86; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 133 TATCATCTGAGTTCCTGTTGCAATGACGAACTTTCCCGAAGAGCTGTAAAT 192
 Db 79127 TACCATCTTGAAGCTTCTTCCAGATGCGCGAGTTCTCCCGATGAGTGAAG 79186

QY 193 AGCTTCAAGTCGTTGGCAGATCTGTCTTGTCTTGATGCAAGATATCCCTGCTTCACTC 252
 Db 79187 GACTTGAAGTCTTGAATATGCTCTTGTCTTGTGCGATGAGAGCCCTTGTCATCGGTA 79246

QY 253 AGTCTCACTTCACCTGCAA 272
 Db 79247 AACCTACTTCACTGAAA 79266

RESULT 9
 AP004061
 LOCUS
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 2 clone
 OJ1407 E09, *** SEQUENCING IN PROGRESS ***.
 ACCESSION AP004061
 VERSION AP004061.1 GI:15208429
 KEYWORDS HTG; PHASE2.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
 clone:OJ1407 E09
 JOURNAL Published Only in Database (2001)
 REFERENCE 2 (bases 1 to 133387)
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 JOURNAL Direct Submission
 Submitted (15-AUG-2001) Takuji Sasaki, National Institute of
 Agrobiological Resources, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rsgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 COMMENT The nucleotide sequence of this BAC clone was generated by
 combining Monsanto and RGP-Japan sequencing data.
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..133387
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="2"
 /clone="OJ1407_E09"

ORIGIN
 Query Match 9.7%; Score 53.6; DB 2; Length 133387;
 Best Local Similarity 61.4%; Pred. No. 0.015;

	Matches	86;	Conservative	0;	Mismatches	54;	Indels	0;	Gaps	0;
QY	133	TATCATCTGAGATTCCCTGTTGGAATTCAGACAACTTTTCCCGAA	CAGCGTCTATAAAAT	192						
Db	132967	TACCATCTCTGAGCTCTTCTTCGAAATGCGCAGATTCTCCCGGATAGGTGA	AAAG	133026						
QY	193	AGCTTCAGTGTGTGGGAACTGTGTCTGTGTGATGAAAGATATCCCTGTG	CATC	252						
Db	133027	GACTTGAGGTCCTTGCAAATCTGCCCTCTGTGTGTCGGATGCGAGCCCTTGT	CATGGTA	133086						
QY	253	AGTCACTTCACCTGC	AA	272						
Db	133087	AACTTACTTCTACCTGAAA	133106							

	RESULT 10			
166494				
LOCUS	166494	7218 bp	DNA	linear
DEFINITION	Sequence 14 from patent US 5670367.			PAT 28-DEC-1997
ACCESSION	166494			
VERSION	166494.1	GI:2724471		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
	Unclassified.			
REFERENCE	1 (bases 1 to 7218)			
AUTHORS	Dorner, F., Schellflinger, F. and Falkner, F. Gunter.			
TITLE	Recombinant fowlpox virus			
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;			
FEATURES	Location/Qualifiers			
source	1..7218			

ORIGIN

Query Match 9.6%; Score 53.2; DB 6; Length 7218;
Best Local Similarity 3.4%; Pred. NO. 0.023;
Matches 13; Conservative 219; Mismatches 152; Indels 0; Gaps 0;

QY	93	CTTTGGCAGCAGCAGCAAACTCCCTTTCTATCTATCTATCTATCTATCTGAGTCTGTG	152
Db	1078	YYY	1137
QY	153	TTGCAAATTCGACAACTTTTCCCGACAGCGCTGTAAATAATAGCTTCAAGTCGTCGCAT	212
Db	1138	YYY	1197
QY	213	CTGTGCTCTGTCTGTGAGCAAGATATCCCTTGCTTCACTCACTCACTTCCACTCGAA	272
Db	1198	YYY	1257
QY	273	GTCATCTAGTTGACCAAAATGAGAACTCTGGGACATCTATATGACGCCCAATPAACTCTT	332
Db	1258	YYY	1317
QY	333	CTCTCCTTGTATGATCATTCCTCCCTTGACTTTTAAGCTCACTCATATGTATACCAAC	392
Db	1318	YYY	1377
QY	393	ACGTTCTGTGTCGAACTATCACCATGAATGCAACATTTTCTAATAATCTGAACCTGA	452
Db	1378	YYGTGA	1437
QY	453	CGAGTACCTTGAAAAAATTATGCT	476
Db	1438	CGAAATTTCTATATCTTTTAAGT	1461

LOCUS	AC074259	163277 bp	DNA linear	HTG 17-JUL-2001
DEFINITION	Trypanosoma brucei chromosome VI clone PC193-4M18, *** SEQUENCING IN PROGRESS ***.			

ACCESSION	AC074259
VERSION	AC074259.13
KEYWORDS	HTG: HTGS PHASE2
SOURCE	Typanosoma brucei
ORGANISM	Typanosoma brucei
REFERENCE	Bukhariyot, Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
AUTHORS	1 (bases 1 to 163277) El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K., Peterson,J., Hou,L., Zhao,H., Mason,T., Miltchev,J., Pal,G., Van Aken,S., Uteerback,T., Khalak,H.G., Gerard,C., Leach,V., Ullu,E., Miyawala,S., White,O., Adams,M.D., Donaldson,J.E. and Fraser,C.M.
TITLE	Typanosoma brucei GUTat10.1 RPi93-4M18 BAC genomic sequence
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 163277)
AUTHORS	El-Sayed,N.M., Khalak,H. and Adams,M.D.
TITLE	Direct Submission
JOURNAL	Submitted (24-JUL-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
COMMENT	On Jul 17, 2001 this sequence version replaced gi:11395701.

FEATURES	Location/Qualifiers
source	1. .163277

ORIGIN

Query Match	Score	DB 2	Length
8.2%	45.2	163277	

Matches 86; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

247 TCACTCAGTCTCACTTCCACCCTGCAAGTCATCTAGTTCACCAATGAGAACTCTGGACA 306

DD 08322 1CACCC1GACACGCGCGC1GGAGG1CG1GGA11CAAGAAHHAGAAHA111CCGA1H 0888

27

[illegible]

RESULTS 12
AE014833

	200608	200709	200810	200911	201012	201101	201202	201303	201404	201505	201606	201707	201808	201909	202010	202111	202212	202301	202402	202503	202604	202705	202806	202907	203008	203109	203210	203311	203412	203501	203602	203703	203804	203905	204006	204107	204208	204309	204410	204511	204612	204701	204802	204903	205004	205105	205206	205307	205408	205509	205610	205711	205812	205901	206002	206103	206204	206305	206406	206507	206608	206709	206810	206911	207012	207101	207202	207303	207404	207505	207606	207707	207808	207909	208010	208111	208212	208301	208402	208503	208604	208705	208806	208907	209008	209109	209210	209311	209412	209501	209602	209703	209804	209905	210006	210107	210208	210309	210410	210511	210612	210701	210802	210903	211004	211105	211206	211307	211408	211509	211610	211711	211812	211901	212002	212103	212204	212305	212406	212507	212608	212709	212810	212911	213012	213101	213202	213303	213404	213505	213606	213707	213808	213909	214010	214111	214212	214301	214402	214503	214604	214705	214806	214907	215008	215109	215210	215311	215412	215501	215602	215703	215804	215905	216006	216107	216208	216309	216410	216511	216612	216701	216802	216903	217004	217105	217206	217307	217408	217509	217610	217711	217812	217901	218002	218103	218204	218305	218406	218507	218608	218709	218810	218911	219012	219101	219202	219303	219404	219505	219606	219707	219808	219909	220010	220111	220212	220301	220402	220503	220604	220705	220806	220907	221008	221109	221210	221311	221412	221501	221602	221703	221804	221905	222006	222107	222208	222309	222410	222511	222612	222701	222802	222903	223004	223105	223206	223307	223408	223509	223610	223711	223812	223901	224002	224103	224204	224305	224406	224507	224608	224709	224810	224911	225012	225101	225202	225303	225404	225505	225606	225707	225808	225909	226010	226111	226212	226301	226402	226503	226604	226705	226806	226907	227008	227109	227210	227311	227412	227501	227602	227703	227804	227905	228006	228107	228208	228309	228410	228511	228612	228701	228802	228903	229004	229105	229206	229307	229408	229509	229610	229711	229812	229901	230002	230103	230204	230305	230406	230507	230608	230709	230810	230911	231012	231101	231202	231303	231404	231505	231606	231707	231808	231909	232010	232111	232212	232301	232402	232503	232604	232705	232806	232907	233008	233109	233210	233311	233412	233501	233602	233703	233804	233905	234006	234107	234208	234309	234410	234511	234612	234701	234802	234903	235004	235105	235206	235307	235408	235509	235610	235711	235812	235901	236002	236103	236204	236305	236406	236507	236608	236709	236810	236911	237012	237101	237202	237303	237404	237505	237606	
--	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--

ACCESSION AE014833 AE014185

KEYWORDS

ORGANISM Plasmodium falciparum 3D7
Eukaryotes: Alveolates: Haemostorida: Plasmodium

REFERENCE	AUTHORS
1 (pages 1 to 252394)	Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W.,

James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A.,

Anginoli, S., Perrea, M., Allen, J., Selengut, J., Haft, D.,
Mather, M.W., Vaidya, A.B., Martin, M.A., Fairhead, A.H.,
Fraunholz, M.J., Koo, D.S., Ralph, S.A., McFadden, G.I.,
Cummins, L.M., Subramanian, G.M., Mungall, C., Venter, J.C.,
Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M. and
Barrell, B.
TITLE
Genome sequence of the human malaria parasite *Plasmodium falciparum*
JOURNAL
Nature 415 (6306), 498-511 (2002)
PUBMED
12368864
REFERENCE
2 (bases 1 to 252394)
AUTHORS
Gardner, M.J.
TITLE
Direct Submission
JOURNAL
Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
FEATURES
Location/Qualifiers

source
1. 252394
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="10"
complement(79. .213)
/note="AT rich"
/rpt_type=tandem
253. .301
repeat_region
/rpt_type=tandem
complement(253. .301)
/rpt_type=tandem
/rpt_unit="(TA)n"
318. .374
repeat_region
/rpt_type=tandem
324. .374
repeat_region
/rpt_type=tandem
/rpt_unit="(TAA)n"
complement(433. .482)
/rpt_type=tandem
/rpt_unit="(TA)n"
577. .605
repeat_region
/rpt_type=tandem
complement(577. .605)
/rpt_type=tandem
/rpt_unit="(A)n"
complement(661. .682)
repeat_region
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(792. .942)
/rpt_type=tandem
/rpt_unit="(TA)n"
897. .1030
repeat_region
/rpt_type=tandem
/rpt_unit="(TATAA)n"
complement(1038. .1173)
/note="AT rich"
/rpt_type=tandem
1100. .1130
repeat_region
/rpt_type=tandem
complement(1100. .1126)
/rpt_type=tandem
/rpt_unit="(TA)n"
1133. .1172
repeat_region
/rpt_type=tandem
1175. .1207
repeat_region
/rpt_type=tandem
complement(1175. .1207)
/rpt_type=tandem
/rpt_unit="(A)n"
complement(1218. .1281)
/note="AT rich"
/rpt_type=tandem
1365. .1391
repeat_region
/rpt_type=tandem
complement(1365. .1391)
/rpt_type=tandem

repeat_region
complement(1401. .1493)
/note="AT rich"
/rpt_type=tandem
1505. .1526
repeat_region
/rpt_type=tandem
/rpt_unit="(A)n"
complement(1602. .1685)
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(1726. .1825)
/note="AT rich"
/rpt_type=tandem
complement(1963. .2007)
/note="AT rich"
/rpt_type=tandem
complement(2049. .2157)
/rpt_type=tandem
/rpt_unit="(TA)n"
2064. .2131
repeat_region
/rpt_type=tandem
complement(2194. .2390)
/rpt_type=tandem
/rpt_unit="(TA)n"
2257. .2332
repeat_region
/rpt_type=tandem
complement(2405. .2457)
/note="AT rich"
/rpt_type=tandem
complement(2481. .2516)
/note="AT rich"
/rpt_type=tandem
2532. .2615
repeat_region
/rpt_type=tandem
/rpt_unit="(A)n"
complement(2623. .2657)
/note="AT rich"
/rpt_type=tandem
2666. .2699
repeat_region
/rpt_type=tandem
/rpt_unit="(TAA)n"
complement(2740. .2807)
/note="AT rich"
/rpt_type=tandem
2855. .3047
repeat_region
/rpt_type=tandem
complement(2856. .3047)
/rpt_type=tandem
/rpt_unit="(TA)n"
2893. .3047
repeat_region
/rpt_type=tandem
3100. .3137
repeat_region
/rpt_type=tandem
complement(3100. .3137)
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(3145. .3170)
/note="AT rich"
/rpt_type=tandem
complement(3196. .3240)
/note="AT rich"
/rpt_type=tandem
3283. .3319
repeat_region
/rpt_type=tandem
complement(3287. .3319)
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(3367. .3434)
/note="AT rich"
/rpt_type=tandem
complement(3513. .3656)
/rpt_type=tandem
/rpt_unit="(TA)n"


```
* 12939 13829: contig of 891 bp in length
* 12930 13829: gap of 100 bp
* 12930 13830 14854: contig of 925 bp in length
* 14855 14954: gap of 100 bp
* 14955 15863: contig of 909 bp in length
* 15864 15963: gap of 100 bp
* 15964 16500: contig of 887 bp in length
* 16501 16950: gap of 100 bp
* 16951 17871: contig of 921 bp in length
* 17872 17971: gap of 100 bp
* 17972 18858: contig of 887 bp in length
* 18859 18959: gap of 100 bp
* 18959 19690: contig of 732 bp in length
* 19691 19790: gap of 100 bp
* 19791 20687: contig of 897 bp in length
* 20688 20787: gap of 100 bp
* 20788 21671: contig of 884 bp in length
* 21672 21771: gap of 100 bp
* 21772 22673: contig of 902 bp in length
* 22674 22773: gap of 100 bp
* 22774 23665: contig of 892 bp in length
* 23666 23765: gap of 100 bp
* 23766 24662: contig of 897 bp in length
* 24663 24762: gap of 100 bp
* 24763 25690: contig of 928 bp in length
* 25691 25790: gap of 100 bp
* 25791 26696: contig of 906 bp in length
* 26697 26796: gap of 100 bp
* 26797 27710: contig of 914 bp in length
* 27711 27810: gap of 100 bp
* 27811 28693: contig of 883 bp in length
* 28694 28793: gap of 100 bp
* 28794 29700: contig of 907 bp in length
* 29701 29800: gap of 100 bp
* 29801 30699: contig of 899 bp in length
* 30700 30799: gap of 100 bp
* 30800 31699: contig of 900 bp in length
* 31700 31799: gap of 100 bp
* 31800 32718: contig of 919 bp in length
* 32719 32818: gap of 100 bp
* 32819 33744: contig of 926 bp in length
* 33745 33844: gap of 100 bp
* 33845 34734: contig of 890 bp in length
* 34735 34834: gap of 100 bp
* 34835 35771: contig of 937 bp in length
* 35772 35871: gap of 100 bp
* 35872 36789: contig of 918 bp in length
* 36790 36889: gap of 100 bp
* 36890 37797: contig of 908 bp in length
* 37798 37897: gap of 100 bp
* 37898 38819: contig of 922 bp in length
* 38820 38919: gap of 100 bp
* 38920 39854: contig of 935 bp in length
* 39855 39954: gap of 100 bp
* 39955 40860: contig of 906 bp in length
* 40861 40960: gap of 100 bp
* 40961 41848: contig of 888 bp in length
* 41849 41948: gap of 100 bp
* 41949 42837: contig of 889 bp in length
* 42838 42937: gap of 100 bp
* 42938 43842: contig of 905 bp in length
* 43843 43942: gap of 100 bp
* 43943 44850: contig of 908 bp in length
* 44851 44950: gap of 100 bp
* 44951 45883: contig of 933 bp in length
* 45884 45983: gap of 100 bp
* 45984 46879: contig of 896 bp in length
* 46880 46979: gap of 100 bp
* 46980 47858: contig of 879 bp in length
* 47859 47958: gap of 100 bp
* 47959 48839: contig of 881 bp in length
* 48840 48939: gap of 100 bp
* 48940 49851: contig of 912 bp in length
```

```
* 49852 49951: gap of 100 bp
* 49952 50886: contig of 935 bp in length
* 50887 50986: gap of 100 bp
* 50987 51916: contig of 930 bp in length
* 51917 52016: gap of 100 bp
* 52017 52912: contig of 836 bp in length
* 52913 53012: gap of 100 bp
* 53013 53911: contig of 899 bp in length
* 53912 54011: gap of 100 bp
* 54012 54890: contig of 879 bp in length
* 54891 54990: gap of 100 bp
* 54991 55881: contig of 891 bp in length
* 55882 55981: gap of 100 bp
* 55982 56858: contig of 877 bp in length
* 56859 56958: gap of 100 bp
* 56959 57860: contig of 902 bp in length
* 57861 57960: gap of 100 bp
* 57961 58859: contig of 899 bp in length
* 58860 58959: gap of 100 bp
* 58960 59855: contig of 896 bp in length
* 59856 59955: gap of 100 bp
* 59956 60872: contig of 917 bp in length
* 60873 60972: gap of 100 bp
* 60972 61865: contig of 893 bp in length
* 61866 61965: gap of 100 bp
* 61966 62867: contig of 902 bp in length
* 62868 62967: gap of 100 bp
* 62968 63861: contig of 894 bp in length
* 63862 63961: gap of 100 bp
* 63962 64858: contig of 887 bp in length
* 64859 64958: gap of 100 bp
* 64959 65851: contig of 893 bp in length
* 65852 65951: gap of 100 bp
* 65952 66849: contig of 898 bp in length
* 66850 66949: gap of 100 bp
* 66950 67838: contig of 885 bp in length
* 67839 67938: gap of 100 bp
* 67939 68873: contig of 935 bp in length
* 68874 68973: gap of 100 bp
* 68974 69883: contig of 910 bp in length
* 69884 69983: gap of 100 bp
* 69984 70890: contig of 907 bp in length
* 70891 70990: gap of 100 bp
* 70991 71930: contig of 940 bp in length
* 71931 72030: gap of 100 bp
```

Query Match 8.1%; Score 45; DB 2; Length 126038;
Best Local Similarity 50.7%; Pred. No. 1.7;
Matches 108; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
QY 282 TTCAACCAATGAGACTCTGGACATCTATATGACCCCAATAAATCTTCTCCTTG 341
Db 95436 TTCCCGAGGTAAGAACCTTTACCAAAAATTTATCCCAAGAAATATCCGACAG 95377
QY 342 TATGATCATTCCTCCCTTTGACTTTAAGTGCACTCATATGATATACCAACAGCTTCTT 401
Db 95376 TATTCAGATATTCATTTAGGTGTCTGCTTCTTAAGTTCCTTGAACAGGTTCTT 95317
QY 402 GTTCGCAATGATGCAATGATGCAATTTCTTAATATATGACGATGACGATACCT 461
Db 95316 GTCCCTTCTTTGATATTTATTCATGAAATTTGATATTTGAATTAAGGTCAGGTTCCCT 95257
QY 462 TGAATAATTTAGCTGAAAAAACCTGATTCTA 494
Db 95256 TCACTAATTTTATTTGAAATATATGTAATTTTA 95224
```

RESULT 14
BV089971/c 564 bp DNA linear STS 15-OCT-2003
LOCUS BV089971
DEFINITION RPAMSE0001867 Roche Palo Alto Mus musculus STS genomic, sequence
tagged site.
ACCESSION BV089971

VERSION BV089971.1 GI:37667450
 KEYWORDS STS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 564)
 AUTHORS Usuka, J., Liao, G., Cheng, J., Nguyen, A., Bach, C., Puech, A., McPherson, J. D., Fournier, D. and Peltz, G.
 TITLE Mus musculus SNPs
 JOURNAL Unpublished (2003)
 COMMENT
 Contact: Jonathan Usuka
 Roche Palo Alto Genetics and Genomics Department
 Roche Palo Alto
 3431 Hillview Ave, Mailstop 83-1, Palo Alto, CA 94024, USA
 Tel: 6508555607
 Email: Jonathan.Usuka@roche.com
 Primer A: No primer submitted
 Primer B: No primer submitted.
 FEATURES
 source
 1..564
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /map="5-8454-8031-CAA01214661.1.1.43014"
 /clone_id="Roche Palo Alto"
 /note="SNPs developed from assay sequences derived from 15 different strains of mice (as of October 1, 2003). Those strains include A/J, A/HeJ, ~129/Sv, AKR/J, B10.D2-H2/Osnu, BALB/cByJ, BALB/cJ, C3H/HeJ, C57BL/6J, CAST/Ei, DBA/2J, MRJ/MpJ, NZB/Blnd, NZW/Iac, SPRET/Ei."

ORIGIN
 STS
 Query Match 7.8%; Score 43.2; DB 11; Length 564;
 Best Local Similarity 63.5%; Pct. No. 6.7;
 Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 40 AAGAGATTGTACAGAAATCAAGTACGACTAAATTAACAAGAAATGATGGCTTTGT 99
 Db 551 AAAGTATTGTACAAATCAATATATAGATATCTCTGGTACCCAGTGGCCCTTTGT 492
 QY 100 CACAGATACGAATATCTCTCTTCTATCTATCTATCTATCTCT 143
 Db 491 CCGGTCCAAATTTTATCTATCTATCTATCTATCTATCTAT 448

RESULT 15
 AC112682 218313 bp DNA linear HTG 05-DEC-2003
 LOCUS Mus musculus chromosome 6 clone RP23-332K10 map 6, *** SEQUENCING
 DEFINITION IN PROGRESS ***, 6 ordered pieces.
 AC112682
 AC112682.5 GI:38708110
 HTG: HTGS_PHASE2; HTGS_FULFILL; HTGS_ACTIVEFIN.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 218313)
 AUTHORS Birren, B., Nussbaum, C. and Lander, E.
 TITLE Mus musculus chromosome 6, clone RP23-332K10
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 218313)
 AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Colliamore, A., Cook, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

TITLE JOURNAL
 REFERENCE AUTHORS
 Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Lenoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norby, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Sudramanian, A., Talmas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (22-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 218313)
 Birren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Colliamore, A., Cook, A., Cooke, P., Corum, B., Dearlano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hago, B., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Liu, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norby, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Sudramanian, A., Talmas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (05-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 5, 2003 this sequence version replaced gi:27884867.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L22646
 Center clone name: 332_K10

NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
 This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
 1 67544: contig of 67544 bp in length
 67545 67644: gap of 100 bp
 67645 132358: contig of 64714 bp in length
 132359 132458: gap of 100 bp
 132459 157227: contig of 24769 bp in length
 157228 157327: gap of 100 bp
 157328 197939: contig of 40612 bp in length
 197940 198039: gap of 100 bp
 198040 203890: contig of 5851 bp in length
 203891 203990: gap of 100 bp

FEATURES * 203991 218313: contig of 14323 bp in length.
 Location/Qualifiers
 1..218313

/organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="6"
 /map="6"
 /clone="RP23-332K10"
 /clone_lib="RP23-332K10 Female Mouse BAC"

ORIGIN

Query Match 7.6% Score 42; DB 2; Length 218313;
 Best Local Similarity 48.7%; Pred. No. 8.4; Indels 0; Gaps 0;
 Matches 114; Conservative 0; Mismatches 120;

QY 119 TCTTTCTATCTATCTATCTATCTGAGTCCCTGTTGCAATGCAACAACCTTTCCCGAA 178
 Db 88582 TCTATCTATCTATCTATCTATCTATCTATCTATCTGCTGCTGCTGCTATTTATCAT 88641
 QY 179 CAGGCTGTAATAATAGCTTCAAGTGTGCGCATCTGTCTGTGTTGATGCAAGATAT 238
 Db 88642 CTATCTATATATATGACCTCTATATTTATCTATATTTATGACTGTTGAGCTAATAT 88701
 QY 239 CCTTGTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 298
 Db 88702 ATATCCATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 88761
 QY 299 CTGGGACATCTAATGACCCCAATAACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 352
 Db 88762 TACCTACCTACCTATCTGTTACTATTAATCACTATTTTATATACACTTT 88815

Search completed: March 19, 2004, 22:27:39
 Job time : 2539 secs

```
OTHER INFORMATION: Description of Artificial Sequence: Consensus
; OTHER INFORMATION: sequence
; NAME/KEY: MOD_RES
; LOCATION: (1)..(2135)
; OTHER INFORMATION: "n" base may be a, t, c or g
US-09-933-711B-17

Query Match
Best Local Similarity 6.0%; Score 33; DB 4; Length 2135;
Matches 64; Conservative 61; Mismatches 110; Indels 0; Gaps 0;

QY 305 CATCAATATGACCCCAATAAATCTCTCTCCCTGTATGATCCATCCCTTGAAT 364
D 1255 CACAAATACACCTCCGAAATTTTGGCAVGNATSAVAYACAGSGMERTGTGCHGT 1314
QY 365 TTAAGTCACTCATATGATATTAACCAACAGCTTCTTGTTCGAACTATCAACATG 424
D 1315 RGATATVOCWYCSHRDGAHYMAAARSAFTYTMVRAITWCTKGYWCAGAYWCMRWYG 1374
QY 425 CAACATTTTCTTAATATCTGAACCTGACGAGTACCTTGAAAAATTTAGCTGAAAA 484
D 1375 CTAATATACGACCAAGGAGATCTTACAGTCTGACGAAATATTAAGATATAA 1434
QY 485 CTTGATTTCTATCCGCAAGCATTTTGTACATCTTTTACTCTGCGCCGCAAG 539
D 1435 AGGDAAGATATGAARATGAGCCWTCMTKTYTGATTSRRRYCAKCKGHR 1489

RESULT 11
US-08-960-048-2
; Sequence 2, Application US/08960048C
; Patent No. 6271443
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; FILE REFERENCE: 15621/01/US
; CURRENT APPLICATION NUMBER: US/08/960,048C
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/029,987
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 4612
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-08-960-048-2

Query Match
Best Local Similarity 6.0%; Score 33; DB 3; Length 4612;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 303 GACATCTATATGACCCCAATAAATCTCTCTCCCTGTATGATCCATCCCTTGAC 362
D 4321 GACACTTAACGACACTTAAGGAGATTTATCTTTGTAAGGCTCCTATGAT 4380
QY 363 TTTTAAGTCACTCATATGATATTAACCAACAGCTTCTTGTTCGAACTATCAATGAA 422
D 4381 TTTGTGTTCATTTATTTCTATATGATTTTCCCTAAATAGAACAAATTTTCA 4440
QY 423 TGCACATTTTCTTAATATCTGAACCTGACGATACCTTGAAAAATTTAGCTGAAAA 482
D 4441 GAGGATATCTAAACCAACAGATTAATTAAGATATACCAAAAAAATATTTCTTAAT 4500
QY 483 AACCT 487
D 4501 TAACT 4505

RESULT 12
```

```
US-09-838-586-2
; Sequence 2, Application US/09838586
; Patent No. 6576818
; GENERAL INFORMATION:
; APPLICANT: Stalker, D. et al.
; TITLE OF INVENTION: Plant Cellulose Synthase and Promoter
; FILE REFERENCE: 15621/02/US
; CURRENT APPLICATION NUMBER: US/09/838,586
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/029,987
; PRIOR FILING DATE: 1996-10-29
; PRIOR APPLICATION NUMBER: 08/960,048
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 4612
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Oligonucleotide
US-09-838-586-2

Query Match
Best Local Similarity 6.0%; Score 33; DB 4; Length 4612;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 303 GACATCTATATGACCCCAATAAATCTCTCTCCCTGTATGATCCATCCCTTGAC 362
D 4321 GACACTTAACGACACTTAAGGAGATTTATCTTTGTAAGGCTCCTATGAT 4380
QY 363 TTTTAAGTCACTCATATGATATTAACCAACAGCTTCTTGTTCGAACTATCAATGAA 422
D 4381 TTTGTGTTCATTTATTTCTATATGATTTTCCCTAAATAGAACAAATTTTCA 4440
QY 423 TGCACATTTTCTTAATATCTGAACCTGACGATACCTTGAAAAATTTAGCTGAAAA 482
D 4441 GAGGATATCTAAACCAACAGATTAATTAAGATATACCAAAAAAATATTTCTTAAT 4500
QY 483 AACCT 487
D 4501 TAACT 4505

RESULT 13
US-09-813-133A-3/C
; Sequence 3, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; FILE REFERENCE: C0001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-3

Query Match
Best Local Similarity 6.0%; Score 33; DB 4; Length 55827;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 168 CTTTCCGACAGGCTGTAAATATAGCTTCAAGTGGCGCATCTGTGTCTTG 227
D 18117 CTGTGAGTTATATGATATGCAAAATATCTTGTAGAGTTGTATATTTGTCTTCTTT 18058
```


QY 228 ATGCAAGATATCCCTTGCTTCATCTCAGTCTTCATCCAGCTGCAAGTCACTAGTACC 287
Db 18057 CTTGATGATGCTGCTTTTCTTTTAACTTCCTTCAGGTTGGTGAATG 17998
QY 288 AATGAGACCTCGGACATCATATAGACCCCAATAACTCTTCTCTCTTATGAT 347
Db 17997 TCTTCTGATATCTGGATATTAAGATATCTCATACTCTTCTGCTCTTATTT 17938
QY 348 CCATCCCTTGACTTTAA 368
Db 17937 AATTTTAATTCATTGAAA 17917
RESULT 14
US-09-790-988-1/c
Sequence 1, Application US/09790988
Patent No. 6632935
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: MATSUNABE, HIDEKI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1
Query Match 6.0%; Score 33; DB 4; Length 640681;
Best Local Similarity 49.2%; Pred. No. 45;
Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 225 TTGATGCAAGATATCCCTTGCTTCATCTCAGTCTTCATCCAGCTGCAAGTCACTAGTTC 284
Db 507116 TTTCTGATGATTTTTCCTGCTTGTTTCATGATCCATCTTACCAACATATATGTTATTT 507057
QY 285 ACCAATGAGACTCTGGGACATCTATATGACCCCAATAACTCTTCTCTGTAAT 344
Db 507056 TGACATTAACACATCTACATGATTTGCTGTGTAATTAATTCACCTAATTTACTAA 506997
QY 345 GATCCATTCCTCTTACTTTTAAAGTCAACTCATATGATATACCAACAGCTTTCTT 401
Db 506996 GTTCTATGCTCTTTTCATGTAATAATATGTTCAATTTTAAACCAACATATTTT 506940
RESULT 15
US-08-916-421B-1/c
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Built et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco-
Patent No. 6503729
TITLE OF INVENTION: jannaschi
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patent In version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschi

FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g

```

NAME/KEY: misc_feature
LOCATION: (559167) ..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241) ..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992) ..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708) ..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081) ..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203) ..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435) ..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442) ..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652) ..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684) ..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455) ..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676) ..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539) ..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619) ..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830) ..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846) ..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881) ..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881) ..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988) ..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224) ..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473) ..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491) ..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091) ..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020) ..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature

```

```

LOCATION: (1602912) ..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734) ..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998) ..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854) ..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

Query Match 6.0%; Score 33; DB 4; Length 1664976;
 Best Local Similarity 46.7%; Pred. NO. 60;
 Matches 105; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

QY 207 GCGCATCTGTGCTCTTCTTGTATGCAAGATATCCCTTGCTTCACTCACTGCTTCCAC 266
DB 1436207 GCAATGTGAATGGGATTTTAAATTTATAGCTTTTAAACAATATCTTGCTTTAT 1436148
QY 267 CTGCAAGTCATCTAGTTCACCAATGAGACTCTGGACATCTATATGACCCCAATAA 326
DB 1436147 CTTTCCATCTCTTCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1436088
QY 327 CTTCTTCTCTCTTGTATGATTCATTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 386
DB 1436087 CTTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1436028
QY 387 ACCAAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 431
DB 1436027 AGACGTATTTTAAATGCTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 1435983

```

Search completed: March 19, 2004, 21:39:31
 Job time : 76 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2004, 20:09:56 ; Search time 2134 Seconds

(without alignments)
7738.419 Million cell updates/sec

Title: US-09-531-113-5981

Perfect score: 553
Sequence: 1 gtaggtatttcagacatc.....ccaaggagaaatgatggag 553

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_estbnum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vtc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pdg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	371.6	67.2	645	AM349814	AM349814 GM210005B
2	323.8	58.6	425	BM270700	BM270700 sak16e09-
3	305.4	55.2	513	BM785857	BM785857 sak8e02-
4	300.8	54.4	383	BG238159	BG238159 sab62h12.

C 5	279.2	50.5	432	10	AM620829	AM620829 sj47a12.Y
C 6	272.8	49.3	375	13	BU964888	BU964888 sat03h07.
C 7	255.8	46.3	576	28	BH405680	BH405680 Gm_UMB001
C 8	228.2	41.3	619	10	BE943027	BE943027 EST422606
C 9	228.2	41.3	637	14	CB891233	CB891233 EST648202
C 10	228.2	41.3	680	12	BM812845	BM812845 EST590938
C 11	228.2	41.3	730	12	BM814531	BM814531 EST592625
C 12	228.2	41.3	758	14	CA922589	CA922589 EST640307
C 13	227.8	40.5	532	28	AO989424	AO989424 Gm_UMB001
C 14	223.8	40.2	712	13	BU548744	BU548744 GMB80015B
C 15	221.2	40.0	726	12	BG586953	BG586953 EST488722
C 16	205.2	37.1	252	12	BG510324	BG510324 sac76d01.
C 17	205.2	37.1	605	12	BM308189	BM308189 sak42g12.
C 18	188.2	34.0	595	14	CA825561	CA825561 R61B04.Tw
C 19	187.6	33.9	498	14	CF606656	CF606656 GEMMA01.0
C 20	186.2	33.7	561	13	BQ623572	BQ623572 USDA-FP.0
C 21	186.2	33.7	590	13	BQ624319	BQ624319 USDA-FP.0
C 22	186.2	33.7	835	14	CB973441	CB973441 CAB30003
C 23	186	33.6	729	14	CF518302	CF518302 CAP0006.T
C 24	186	33.6	757	14	CB973510	CB973510 CAB30003.T
C 25	186	33.6	769	14	CF519158	CF519158 CAP0006.T
C 26	186	33.6	887	14	CF355583	CF355583 V-B-129G1
C 27	185.2	33.5	765	14	CB347779	CB347779 CAB2SG000
C 28	184.4	33.3	777	14	CF372668	CF372668 GSEGS079A
C 29	178.2	32.2	717	14	CF518533	CF518533 CAP0007.T
C 30	175.8	31.8	747	14	CF231529	CF231529 PAC0022B
C 31	173	31.3	666	13	BQ121429	BQ121429 EST607005
C 32	169.6	30.7	671	13	BQ414824	BQ414824 GA_Ed009
C 33	169.6	30.7	671	13	BQ406165	BQ406165 GA_Ed009
C 34	165	29.8	698	14	CF518449	CF518449 CAP0007.T
C 35	158.2	28.6	571	14	CB347637	CB347637 CAB2SG000
C 36	153.8	27.8	549	10	BE037501	BE037501 MF02D05.M
C 37	153.8	27.8	549	12	BG524880	BG524880 8-57.Stev
C 38	151.2	27.3	634	13	BQ991598	BQ991598 QGF23E22.
C 39	146.6	26.5	749	14	CD846972	CD846972 DHOA53ZA
C 40	145.8	26.4	391	9	AI443193	AI443193 sa45f06.Y
C 41	144.6	26.1	552	14	CB347551	CB347551 CAB2SG000
C 42	142.6	25.8	597	12	BI405821	BI405821 146E08.MA
C 43	142.6	25.8	618	13	BQ121428	BQ121428 EST607004
C 44	140.8	25.5	144	14	CD413174	CD413174 Gm_CK4458
C 45	138.6	25.1	658	13	BQ987454	BQ987454 QGF12E07.

ALIGNMENTS

RESULT 1
AM349814
LOCUS
DEFINITION
AM349814
ACCESSION
AM349814.1 GI:6847453
VERSION
AM349814.1
SOURCE
Glycine max (soybean)
ORGANISM
BST.
Glycine max (soybean)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 645)
Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Corryell,V.,
Expeiding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L., and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other ESTs: A144567
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147

Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Resgen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: ccu@resgen.com
 High quality sequence stop: 366.
 Location/Qualifiers
 1..383

FEATURES

source

/organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1043-4199"
 /issue_type="Hypocotyl and plumule, germinating seeds"
 /lab_host="DH10B"
 /clone_lib="Gm-c1043"
 /note="Vector: pT73Pac (Pharmacia); Site_1: EcoRI;
 Site_2: NotI; This cDNA library was constructed from mRNA
 isolated from hypocotyl and plumule tissues of seeds
 germinated for three days of the cultivar Williams.
 Complementary DNA was synthesized from mRNA using a primer
 consisting of a poly(dT) sequence with a NotI restriction
 site. EcoRI adapters were ligated to the blunt-ended cDNA
 fragments followed by digestion with EcoRI and NotI. The
 cDNA fragments were directionally cloned into the
 EcoRI-NotI restriction site of the pT73-Pac vector. The
 ligated cDNA fragments were transformed into DH10B host
 cells (Gibco BRL). This library was constructed by Dr.
 Randy Shoemaker."

ORIGIN

Query Match 54.4%; Score 300.8; DB 12; Length 383;
 Best Local Similarity 99.3%; Pred. No. 1.2e-70;
 Matches 302; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 GTAGCTATTTCGACGATTACCACTTATTTCGGGAAGAGATTGTACTAGAAATTC 60
DB 304 GTTAGTTAGATCCAGCAATTAAACAATTATTCGGGAAGAGATTGTACTAGAAATTC 245
QY 61 AAGTAGGAGCTAAATATACAAAACCTGATGGCTTTTGTGCACAGATGAGATATCTTCCTC 120
DB 244 AAGTAGGAGCTAAATATACAAAACCTGATGGCTTTTGTGCACAGATGAGATATCTTCCTC 185
QY 121 TTTCATCTATCTATCTATCTCTGAGTTCCTGTTGCAATGACAACTTTCCGGAACA 180
DB 184 TTTCATCTATCTATCTCTGAGTTCCTGTTGCAATGACAACTTTCCGGAACA 125
QY 181 GGCTGTAAATAATCTCAAGCTGCTGGGCACTGCTGCTCTTGTGATGCAAGATATCC 240
DB 124 GGCTGTAAATAATCTCAAGCTGCTGGGCACTGCTGCTCTTGTGATGCAAGATATCC 65
QY 241 CTGCTCTCACTCACTCACTTCCACCTGCAAGTCATCTAGTTCACCAATAGAGACTCT 300
DB 64 CTGCTCTCACTCACTCACTTCCACCTGCAAGTCATCTAGTTCACCAATAGAGACTCT 5
QY 301 GGGA 304
DB 4 GGGA 1

```

RESULT 5
 AM620829/c 432 bp mRNA linear EST 03-DEC-2001
 LOCUS sj47a12.y1 Gm-c1033 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION Gm-c1033-407 5', mRNA sequence.
 AM620829
 ACCESSION AM620829.1 GI:7327031
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Glycine max (soybean)
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

REFERENCE

AUTHORS

1 (bases 1 to 432)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelidng,J., Coryell,V.,
 Khanna,A., Bolla,M., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
 Beck,C., Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M.,
 Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
 Schurr,R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@waton.wustl.edu
 This clone is available through: Resgen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: ccu@resgen.com
 Insert Length: 649 Std Error: 0.00
 High quality sequence stop: 414.
 Location/Qualifiers

FEATURES

source

1..432
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1033-407"
 /issue_type="Desloy 5710' seedling roots"
 /lab_host="DH10B"
 /clone_lib="Gm-c1033"
 /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
 XhoI; This cDNA library was constructed from mRNA isolated
 from 'Desloy 5710' seedling roots. Tissue was taken from
 7-day-old seedlings that had been propagated on paper
 towels with distilled water. Tissue was taken from the tip
 to the first lateral root, usually about 3cm from the tip,
 and flash-frozen in liquid nitrogen. StrataGene's cDNA
 Synthesis Kit (catalog number 200401) was used to
 synthesize the cDNA. First-strand synthesis was performed
 with 5-methyl dCTP, hence the ligated cDNA was
 hemimethylated. A modification of StrataGene's
 first-strand synthesis primer was used. An 'anchor'
 nucleotide (V=A, C, or G) was added to the 3' end of the
 primer [GAGAGAGAGAGAGAGAGAGACTCTCGAG(T)18V] to anchor
 the primer at the 5' end of the poly(A) tract. After
 second strand synthesis, the cDNA ends were filled in with
 cloned Pfu DNA, ligated to EcoRI adapters and
 subsequently phosphorylated. The cDNA was then
 precipitated and redissolved in sterile, RNase-,
 DNase-free water. The XhoI site within the first-strand
 synthesis primer was then restricted by digestion with
 XhoI from Promega (400U/ul); all XhoI sites in the cDNA
 would be protected by their hemimethylated status. The
 cDNA constructs were size-fractionated with a 500 bp
 cutoff, using GabcobRL Life Technologies' cDNA size
 fractionation column. The column eluent was then
 precipitated, redissolved, and ligated into StrataGene's
 pBluescript II XR Predigested vector (pBluescript II SK(+)
 vector that has been digested with EcoRI and XhoI, and
 phosphorylated by StrataGene). 100% of the white and blue
 colonies appear to contain recombinant plasmids with cDNA
 inserts, based on size (n=24 and 6, respectively). This
 library was constructed by Dr. Paul Keim and Dr. Virginia
 Coryell."

ORIGIN

Query Match 50.5%; Score 279.2; DB 10; Length 432;
 Best Local Similarity 96.5%; Pred. No. 8.5e-65;
 Matches 299; Conservative 0; Mismatches 3; Indels 8; Gaps 1;

```

QY 127 TCATCTATCTATCTGAGTTCCTGTTGCAATTGACGAACTTTCCGGAAGAGCTGT 166
DB 432 TCATCTATCTATCTGAGTTCCTGTTGCAATTGACGAACTTTCCGGAAGAGCTGT 373

```

QY 187 AAAATAGCTTCAAGTCTGTCGCGATCTGTCTTCTTGTGATGCAAGATATCCCTTGT 246
 Db 372 AAACATAGCTTCAAGTCTGTCGCGATCTGTCTTCTTGTGATGCAAGATATCCCTTGT 313
 QY 247 TCAGTCAAGTCTCAAGTCTGTCGCGATCTGTCTTCTTGTGATGCAAGATATCCCTTGT 306
 Db 312 TCAGTCAAGTCTCAAGTCTGTCGCGATCTGTCTTCTTGTGATGCAAGATATCCCTTGT 253
 QY 307 TCATATGACCCCAATTAACCTTCTCTCTCTTGTATGATGATTC-----CCCTT 358
 Db 252 TCATATGACCCCAATTAACCTTCTCTCTCTTGTATGATGATTC-----CCCTT 193
 QY 359 TCAGTCAAGTCTCAAGTCTGTCGCGATCTGTCTTCTTGTGATGCAAGATATCCCTTGT 418
 Db 192 TCAGTCAAGTCTCAAGTCTGTCGCGATCTGTCTTCTTGTGATGCAAGATATCCCTTGT 133
 QY 419 TGAATGCAAC 428
 Db 132 TGAATGCAAC 123

RESULT 6 375 bp mRNA linear EST 21-OCT-2002
 LOCUS BU964888/c
 DEFINITION Sat03h07.y1 Gm-c1036 Glycine max cDNA clone SOYBEAN CLONE ID:
 BU964888
 ACCESSION BU964888.1 GI:24205635
 VERSION EST.
 KEYWORDS
 SOURCE
 ORGANISM

Glycine max (soybean)
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 375)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V.,
 Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
 Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
 Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
 Schurr, R., Ritzer, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
 McGinn, R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Reggen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact: cdueregen.com web site:
 www.reggen.com

FEATURES

source
 1..375
 Location/Qualifiers
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-c1036-12662"
 /tissue_type="somatic embryos cultured on MSD 20"
 /lab_host="DH10B"
 /clone_id="Gm-c1036"
 /note="Vector: pSPORT1, Site 1: NotI, Site 2: SalI. This
 cDNA library was constructed from mRNA isolated from
 somatic embryos (age ranging from 2 months to 9 months)
 cultured on MSD 20. The library was prepared using the
 Life Technologies superscript cDNA library construction

kit. Complementary DNA was synthesized from mRNA using a
 poly (dt) sequence with a NotI restriction site. SalI
 linkers adapters were ligated to the blunt-ended cDNA
 fragments followed by NotI digestion. The cDNA fragments
 were directionally cloned into the NotI-SalI restriction
 site of the pSPORT1 vector. The ligated cDNA fragments
 were transformed into E. coli Electromax DH10B host cells.
 This library was constructed in the laboratory of Dr. Lila
 Vodkin by Anu Khanna at the University of Illinois at
 Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

Query Match 49.3%; Score 272.8; DB 13; Length 375;
 Best Local Similarity 99.3%; Pred. No. 4.5e-63;
 Matches 274; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTAGTATATTTCCAGACATTAACACAACTTTATTTCCGGAAGATGTAAGTAATAC 60
 Db 276 GTAGTATATTTCCAGACATTAACACAACTTTATTTCCGGAAGATGTAAGTAATAC 217
 QY 61 AAGTAGCCACTAAATTAACAAACCTGATGCTTTTGTACAGATAGCAATACCTTC 120
 Db 216 AAGTAGCCACTAAATTAACAAACCTGATGCTTTTGTACAGATAGCAATACCTTC 157
 QY 121 TTTTATCTATCTATCTATCTGATGCTTCTGTTGGAATGCAAGCACTTTCCGACA 180
 Db 156 TTTTATCTATCTATCTATCTGATGCTTCTGTTGGAATGCAAGCACTTTCCGACA 97
 QY 181 GCGCTGTAATAATGATCTTCAAGTCTGTTGGCATCTGTCTTCTTGTATGCAAGATATCC 240
 Db 96 GCGCTGTAATAATGATCTTCAAGTCTGTTGGCATCTGTCTTCTTGTATGCAAGATATCC 37
 QY 241 CTGGCTTCACTCAAGTCTCACTCCACGTCGAAGTCA 276
 Db 36 CTGGCTTCACTCAAGTCTCACTCCACGTCGAAGTCA 1

RESULT 7 576 bp DNA linear GSS 04-DEC-2001
 LOCUS BH405680
 DEFINITION Gm_UMB001_109_J13 R UMN soybean BAC library (pCBSBAC ECKR1)
 Glycine max genomic clone Glycine max genomic clone
 Gm_UMB001_109_J13, genomic survey sequence.
 ACCESSION BH405680.1 GI:17351835
 VERSION GSS.
 KEYWORDS
 SOURCE
 ORGANISM
 Glycine max (soybean)
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 576)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Marek, J.F., Paz, M., Darnielle, L., Hanson, N. and Shoemaker, R.C.
 BAC End sequences from a soybean genomic library (ISU)
 Unpublished (2000)
 Contact: Shoemaker Randy C
 Agronomy Department
 Iowa State University
 Ames, IA 50011-1010, USA
 Tel: 515 294 1205
 Fax: 515 294 2299
 Email: rcshoe@iastate.edu
 This BAC identified by SSR Sat429. For more information, see
 SoyBase at:
 http://genome.cornell.edu/cgi-bin/WebPage/webPage?db=soybase. Please
 see as an authority for the mapping/finishing: Cregan P.B., T. Jarvik,
 A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T.
 Vantrolat, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a. An
 integrated genetic linkage map of the soybean genome. Crop Sci.
 39:1464-1490.

Insert Length: 125 Std Error: 0.00
 Seq primer: M13R

Class: BAC ends.
Location/Qualifiers
Source

1.576
/organism="Glycine max"
/mol_type="genomic DNA"
/cultivar="Farldault"
/db_xref="taxon:3847"
/clone="Gm Umb001.109 J13"
/issue_type="cotyledon leaves"
/dev_stage="cotyledon"
/clone_1lb="UMN Soybean BAC library (pECsBAC4 EcORI)
Glycine max genomic clone"
/note="Vector: pECsBAC4; The UMN BAC library (Danesh et al, Theor. Appl. Genet. 96:196, 1998) was constructed using the Eco RI site of pECsBAC4. The library consists of 72,960 clones with an average insert size of 120 kb, equal to 7 haploid genome equivalents. Screening of the library is done by hybridization of high-density colony filters and/or PCR amplification of DNA pools. Four high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening"

ORIGIN

Query Match 46.3%; Score 255.8; DB 28; Length 576;
Best Local Similarity 97.3%; Pred. No. 1.9e-58;
Matches 257; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GTAGGTTATTTCCAGCATTAACAACAATTATTTGGGAGAGATTGTACTGAAATAC 60
DB 285 GTAGGTTAATCCAGCATTAACAACAATTATTTGGGAGAGATTGTACTGAAATAC 344
QY 61 AAGTACGCACTAAATTAACAACAAGTGGCTTTGTCACAGATGAACTCTCC 120
DB 345 AAGTACGCACTAAATTAACAACAAGTGGCTTTGTCACAGATGAACTCTCC 404
QY 121 TTTCTATCTATCTATCTATCTGAGTCTGTCGAATGACGAACCTTTCCGAGAC 180
DB 405 TTTCTATCTATCTATCTATCTGAGTCTGTCGAATGACGAACCTTTCCGAGAC 464
QY 181 GGCTGTAATAATGCTTCAAGTCGTGGCATCTGTCTGTCGTGATGCAAGATATCC 240
DB 465 GGCTGTAATAATGCTTCAAGTCGTGGCATCTGTCTGTCGTGATGCAAGATATCC 524
QY 241 CTGGCTTCACTAGTCTCACTCC 264
DB 525 CTGGCTTCACTAGTCTCACTCC 548

RESULT 8
BE943027/c 619 bp mRNA linear EST 03-OCT-2000
LOCUS EST422606 MGHG Medicago truncatula cDNA clone pmGHG-14H1, mRNA
DEFINITION
sequence.

ACCESSION BE943027
VERSION BE943027.1 GI:10520786
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 619)
AUTHORS Cote, F., Ojane-Reuthe, T., Hahn, M. G., Vandenberg, K., Hur, J., Beremand, P., Andre, G., Town, C. D., Bowman, C. L., Craven, M. B. and Cho, J., Fraser, C. M.
TITLE ESTs from seedling roots of Medicago truncatula after treatment with beta glucan elicitor preparation from Phytophthora sojae

JOURNAL Unpublished (2000)
COMMENT Contact: Michael G. Hahn
Complex Carbohydrate Research Center
University of Georgia

220 Riverbend Road, Athens, GA 30602-4712, USA
Tel: 706-542-4457
Fax: 706-542-4412
Email: hahn@ccrc.uga.edu
University of Georgia name: G270879e TIGR sequence name: MTUBE37TK
More information is available at: <http://chryslite.tamu.edu/medicago>
Seq primer: SKmod (CTA GAA CTA gtc gat cc).
Location/Qualifiers
1.619
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="A17"
/db_xref="taxon:3880"
/clone="pmGHG-14H1"
/issue_type="roots from four day old seedlings"
/dev_stage="2 days after treatment with beta glucan elicitor preparation from Phytophthora sojae"
/lab_host="E. coli strain X10LR"
/clone_1lb="MGHG"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in X10LR cells."

FEATURES

source

ORIGIN

Query Match 41.3%; Score 228.2; DB 10; Length 619;
Best Local Similarity 85.5%; Pred. No. 5.6e-51;
Matches 254; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 132 CATATCATCTCGAGTCTGCTGCAATTCGACGAACCTTTCCGACAGCGTGTAATA 191
DB 389 CATATCATCTGAGTCTGCTGCAATTCGACGAACCTTTCCGACAGCGTGTAATA 330
QY 192 TAGCTTCAAGTGGTGGCATCTGTCTGTCGTGATGACAGATATCCCTGCTTCACT 251
DB 329 TAGCTTCAAGTGGTGGCATCTGTCTGTCGTGATGACAGATATCCCTGCTTCACT 270
QY 252 CAGTCTCACTTCACCTGCACTCATCTAGTTCACCAATGAGAACTGTGGACATCTAT 311
DB 269 AAGTCTCACTTCACCTGCACTCATCTAGTTCACCAATGAGAACTGTGGACATCTAT 210
QY 312 AAGACCCCAATAAATCTTCTCTGCTGTAATGATTCATTCCTCTGACTTTAAGCT 371
DB 209 AAGACCCCAATAAATCTTCTCTGCTGTAATGATTCATTCCTCTGACTTTAAGGA 150
QY 372 CAATCATATGATATACCAACAGTCTTCTGTCGAACTATCACCATGATGCAAC 428
DB 149 CAATCATATGATATACCAACAGTCTTCTGTCGAACTATCACCATGATGCAAC 93

RESULT 9
CB891233/c 637 bp mRNA linear EST 24-APR-2003
LOCUS EST48202 KV3 Medicago truncatula cDNA clone KV3-49H9, mRNA
DEFINITION
sequence.

ACCESSION CB891233
VERSION CB891233.1 GI:30098400
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 637)
AUTHORS Vandenberg, K., Andre, G., Hur, J., Beremand, P., Town, C. D., Van Aken, S., Utterback, T., Cho, J. and Fraser, C. M.
TITLE ESTs from roots of Medicago truncatula 72 h after Rhizobium inoculation, 2001

JOURNAL Unpublished (2001)

JOURNAL

Unpublished (2002)
Other ESTs: ES1592624

Contact: Bird, DM
Plant Nematode Genetics Group
North Carolina State University
Box 7616, Raleigh, NC 27695, USA
Tel: 919 515 6813
Fax: 919 515 9500

Email: david.bird@ncsu.edu
TIGR sequence name: MT04U481KB More information is available at:
www.medicago.org

Seq primer: SKmod (CTA gga cta gtc gat cc).

FEATURES

Source

Location/Qualifiers
1..730
/organism="Medicago truncatula/Meloidogyne incognita mixed
EST library"
/mol_type="mRNA"
/culti_var="A17"
/db_xref="taxon:188702"
/clone="PBNIR-22G23"
/issue_type="3 week old roots"
/dev_stage="3 days after infection with Meloidogyne
incognita second stage larvae"
/lab_host="XIOIR"
/clone_1ib="BNIR"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the UniZap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XIOIR cells."

ORIGIN

Query Match

Best Local Similarity 41.3%; Score 228.2; DB 12; Length 730;
Matches 254; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 132 CTATCATCTCTGAGTTCCTGTTGCAATTGACGAACTTTCCGAAACAGGCTGTAAAAA 191
DB 697 CTATCATCTCTGAGTTCCTGTTGCAATTGACGAACTTTCCGAAACAGGCTGTAAAAA 638
QY 192 TAGCTTCAAGTGGTGGGATCTGTGCTGTTGATGACAAATATCCCTTGGCTTCACT 251
DB 637 TAGCTTCAAGTGGTGGGATCTGTGCTGTTGATGACAAATATCCCTTGGCTTCACT 578
QY 252 CAGTTCACCTTCACCTGCAAGTCATCTAGTTCACCAAAATGAGAACTCTGGACATCTAT 311
DB 577 AAGTTCACCTTCACCTGCAAGTCATCTAGTTCACCAAAATGAGAACTCTGGACATCTAT 518
QY 312 ATGACCCCCCAATAACTTTCTCTCTGTTGATGATCCATTCCTCTTGAATTTAAGGT 371
DB 517 ATGACCCCCCAATAACTTTCTCTCTGTTGATGATCCATTCCTCTTGAATTTAAGGT 458
QY 372 CAAGCTATATGTATATACCAACAGTTCCTGTTCCGAATCATATACCATGATGACAC 428
DB 457 CAAGCTATATGTATATACCAACAGTTCCTGTTCCGAATCATATACCATGATGACAC 401

RESULT 12
LOCUS CA922589 758 bp mRNA linear EST 09-MAY-2003
DEFINITION EST640307 MTUS Medicago truncatula cDNA clone MTUS-55G1, mRNA
sequence.

ACCESSION CA922589
VERSION CA922589.1 GI:27409519
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE

1 (bases 1 to 758)

Vandenbosch, K., Endre, G., Silverstein, K., Town, C.D., Van Aken, S.,
Utterback, F., Cheung, F., and Fraser, C.M.
The Medicago truncatula 6K unigene set: cDNA clones selected and
re-arranged from various libraries

JOURNAL

Unpublished (2002)

Contact: Vandenbosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738

Email: kvandenb@cbbs.umn.edu
Alias Clone pMRAM-47E20

TIGR sequence name: MTUCC731V
More information is available at: www.medicago.org
Seq primer: (gca ata cga ctc act ata ggg c).

FEATURES

Source

Location/Qualifiers
1..758
/organism="Medicago truncatula"
/mol_type="mRNA"
/culti_var="A17"
/db_xref="taxon:3880"
/clone="MTUS-55G1"
/issue_type="mixed libraries"
/dev_stage="various stages"
/lab_host="XIOIR"
/clone_1ib="MTUS"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the UniZap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XIOIR cells."

ORIGIN

Query Match

Best Local Similarity 41.3%; Score 228.2; DB 14; Length 758;
Matches 254; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 132 CTATCATCTCTGAGTTCCTGTTGCAATTGACGAACTTTCCGAAACAGGCTGTAAAAA 191
DB 448 CTATCATCTCTGAGTTCCTGTTGCAATTGACGAACTTTCCGAAACAGGCTGTAAAAA 507
QY 192 TAGCTTCAAGTGGTGGGATCTGTGCTGTTGATGACAAATATCCCTTGGCTTCACT 251
DB 508 TAGCTTCAAGTGGTGGGATCTGTGCTGTTGATGACAAATATCCCTTGGCTTCACT 567
QY 252 CAGTTCACCTTCACCTGCAAGTCATCTAGTTCACCAAAATGAGAACTCTGGACATCTAT 311
DB 568 AAGTTCACCTTCACCTGCAAGTCATCTAGTTCACCAAAATGAGAACTCTGGACATCTAT 627
QY 312 ATGACCCCCCAATAACTTTCTCTCTGTTGATGATCCATTCCTCTTGAATTTAAGGT 371
DB 628 ATGACCCCCCAATAACTTTCTCTCTGTTGATGATCCATTCCTCTTGAATTTAAGGT 687
QY 372 CAAGCTATATGTATATACCAACAGTTCCTGTTCCGAATCATATACCATGATGACAC 428
DB 688 CAAGCTATATGTATATACCAACAGTTCCTGTTCCGAATCATATACCATGATGACAC 744

RESULT 13

LOCUS

AQ989424

DEFINITION

Gm UMD001.154 B19F UMN Soybean BAC Library (PESBAC4 BAC) Glycine
max genomic clone Gm UMD001.154_B19; genomic survey sequence.

ACCESSION

AQ989424

VERSION

AQ989424.1 GI:6937370

KEYWORDS

GSS.

SOURCE

Glycine max (soybean)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 532)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Marek, L.F., Paz, M., Darnielle, L., Hanson, N. and Shoemaker, R. C.
BAC End sequences from a soybean genomic library (ISU)
Unpublished (2000)
Contact: Shoemaker Randy C
Agronomy Department
Iowa State University
Ames, IA 50011-1010, USA
Tel: 515 294 1205
Fax: 515 294 2299
Email: rcschoe@iastate.edu
This BAC identified by SSR Sat429. For more information, see Soybase at:
http://genome.cornell.edu/cgi-bin/WebPage/webase?db=soybase.
Please see as an authority for the mapping/naming: Cregan P. B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. Vantol, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a. An integrated genetic linkage map of the soybean genome. Crop Sci. 39: (in press)
Seq primer: M13P
Class: BAC ends.

FEATURES
SOURCE

Location/Qualifiers
1..532
/organism="Glycine max"
/mol type="genomic DNA"
/cultivar="Faribault"
/db_xref="taxon:3847"
/clone="Gm-UMB001_154_B19;"
/tissue type="cotyledon leaves"
/dev stage="cotyledon stage"
/clone.lib="UMN Soybean BAC Library (PECSBAC4 Ecotri)"
/note="Vector: pECSBAC4; The UMN BAC Library (Darnesh et al; Theor. Appl. Genet. 96:196, 1998) was constructed using the Eco RI site of pECSBAC4. The library consists of 72,960 clones with an average insert size of 120 kb, equal to 7 haploid genome equivalents. Screening of the library is done by PCR amplification of DNA pools."

ORIGIN

Query Match 41.2%; Score 227.8; DB 28; Length 532;
Best Local Similarity 94.6%; Pred. No. 7.1e-51;
Matches 229; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GTAGGTAATTCGAGACATTAACGACATTTTCGGGAAGATGTACTAGAAATAC 60
DB 291 GTAGGTAGATCCGACATTAACCACTTATTTGGGAAAGATGTACTAGAAATAC 350
QY 61 AAGTAGGACCTAAATTAACAAATCTGATGGCTTTTGTCAAGATACGAATCTTCTC 120
DB 351 AAGTAGGACCTAAATTAACAAATCTGATGGCTTTTGTCAAGATACGAATCTTCTC 410
QY 121 TTTCATATCTATCTATCTCTGAGTTCTGTTTGGAAATGACGAACCTTTCCGACA 180
DB 411 TTTCATATCTATCTATCTCTGAGTTCTGTTTGGAAATGACGAACCTTTCCGACA 470
QY 181 GGCGTAAATATAGCTCAAGTGGTGGCATCTGCTCTTGTGATGCAAGATATCC 240
DB 471 GGCGTAAATATAGCTCAAGTGGTGGCATCTGCTCTTGTGATGCAAGATATCC 530
QY 241 CT 242
DB 531 CT 532

RESULT 14
LOCUS BUS48744 712 bp mRNA linear EST 16-SEP-2002
DEFINITION GM880015B20G10 Gm-r1088 Glycine max cDNA clone Gm-r1088-5708 3',
mRNA sequence.
ACCESSION BUS48744

VERSION
KEYWORDS
SOURCE
ORGANISM

BUS48744.1 GI:22931605
EST.
Glycine max (soybean)
Glycine max

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 712)
Vodkin, L., Shoemaker, R., Keim, P., Retzel, E., Khanna, A., Shealy, R., Clough, S., Thibaud-Nissen, F., Corryell, V., Erpelting, J., Rapp, C., Shoop, E., Stromvik, M., Schweitzer, P., Gong, G. and Liu, L.
A Functional Genomics Program for Soybean (NSF 9872565) (2002)
Unpublished (2002)
Other ESTs: BF070750 corresponding to Gm-cl065-2303 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
Insert Length: 712 Std Error: 0.00
Plate: GM880015B20 row: G column: 10
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'
High quality sequence stop: 712.

FEATURES
SOURCE

Location/Qualifiers
1..712
/organism="Glycine max"
/mol type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1088-5708"
/clone.lib="Gm-r1088"
/note="The library Gm-r1088 is a sequence-driven, reracked set of 9,216 low redundancy clones selected from cDNA libraries from various tissues and stages of development of soybean that consists of 2,706 cDNAs from germinating cotyledons (source library Gm-cl027); 1,355 cDNAs from immature seed coats (libraries Gm-cl019, and Gm-cl023); 917 cDNAs from tissue culture derived somatic embryos (source libraries Gm-cl036 and Gm-cl075); 3,275 cDNAs from germinating seedlings, shoot tips, or leaves exposed to various stresses (source libraries Gm-cl064, Gm-cl065, Gm-cl066, and Gm-cl067; and Gm-cl068); and 963 cDNAs from young leaves exposed to bacterial and fungal pathogens (source libraries Gm-cl072, Gm-cl073, and Gm-cl074). The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were reracked to form library Gm-r1088 and the cDNA clones of the reracked Gm-r1088 library were then sequenced at the 3' end. The unique selection and 3' sequencing was funded by NSF Plant Genome Project #9872565 (http://soybeanomics.croplsc.uiuc.edu/) as part of creation of a low redundancy soybean cDNA set. The source cDNA libraries were constructed by the laboratories of Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa State University, and Paul Keim, Northern Arizona University as part of the Public EST project, http://119.186.26.94/soybeanest.htm. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota.
http://web.anc.uiuc.edu/biodata/nfsy/. Reracking and 3' sequencing were conducted by services of the University of Illinois Keck Center for Comparative and Functional Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The corresponding 5' EST from each clone in the Gm-r1088 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER

[illegible]

```

Db      1378 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGGTA 1437
Cy      453 CGAGTACTGTGAAAAATTAGCT 476
        |||||
Db      1438 CCAATTCTTCTACTCTTAACT 1461

RESULT 3
US-09-539-333D-129/c
: Sequence 129, Application US/09539333D
: Patent No. 6476208
: GENERAL INFORMATION:
: APPLICANT: Cohen, Daniel
: APPLICANT: Blumenfeld, Martha
: APPLICANT: Chumakov, Ilya
: APPLICANT: Bougueleret, Lydie
: APPLICANT: Bihain, Bernard
: APPLICANT: Essiloux, Laurent
: TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
: FILE REFERENCE: GENSET .047AUS
: CURRENT APPLICATION NUMBER: US/09/539,333D
: CURRENT FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: US 60/126,903
: PRIOR FILING DATE: 1999-03-30
: PRIOR APPLICATION NUMBER: US 60/131,971
: PRIOR FILING DATE: 1999-04-30
: PRIOR APPLICATION NUMBER: US 60/132,065
: PRIOR FILING DATE: 1999-04-30
: PRIOR APPLICATION NUMBER: US 60/143,928
: PRIOR FILING DATE: 1999-07-14
: PRIOR APPLICATION NUMBER: US 60/145,915
: PRIOR FILING DATE: 1999-07-27
: PRIOR APPLICATION NUMBER: US 60/146,453
: PRIOR FILING DATE: 1999-07-29
: PRIOR APPLICATION NUMBER: US 60/146,452
: PRIOR FILING DATE: 1999-07-29
: PRIOR APPLICATION NUMBER: US 60/162,288
: PRIOR FILING DATE: 1999-10-28
: PRIOR APPLICATION NUMBER: US 09/416,384
: PRIOR FILING DATE: 1999-10-12
: NUMBER OF SEQ ID NOS: 231
: SOFTWARE: Patent.pm
: SEQ ID NO 129
: LENGTH: 3001
: TYPE: DNA
: ORGANISM: Homo Sapiens
: FEATURE:
: NAME/KEY: allele
: LOCATION: 1501
: OTHER INFORMATION: 99-25966-241 : polymorphic base T or C
: FEATURE:
: NAME/KEY: misc_binding
: LOCATION: 1502..1520
: OTHER INFORMATION: 99-25966-241.misl, complement
: FEATURE:
: NAME/KEY: misc_binding
: LOCATION: 1481..1500
: OTHER INFORMATION: 99-25966-241.mis2,
: FEATURE:
: NAME/KEY: primer_bind
: LOCATION: 1721..1741
: OTHER INFORMATION: upstream amplification primer, complement
: FEATURE:
: NAME/KEY: primer_bind
: LOCATION: 1219..1239
: OTHER INFORMATION: downstream amplification primer
: FEATURE:
: NAME/KEY: misc_binding
: LOCATION: 1489..1513
: OTHER INFORMATION: 99-25966-241 probe
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1659,1987

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2004, 20:20:31 ; Search time 72 Seconds
(without alignments)
4262.331 Million cell updates/sec

Title: US-09-531-113-5981

Perfect score: 553
Sequence: 1 gtaggttattccagacatt.....ccaaggagactgacgag 553

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCITUS COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	98.6	17.8	277	4	US-09-313-294A-1048
C 2	53.2	9.6	7218	1	US-08-232-463-14
C 3	35.8	6.5	3001	4	US-09-539-333D-129
C 4	34.4	6.2	580073	4	US-08-545-528D-1
C 5	33.8	6.1	1500	4	US-09-134-001C-2533
C 6	33.6	6.1	246	4	US-09-280-116-144
C 7	33.4	6.0	193303	4	US-09-497-855A-37
C 8	33.4	6.0	193303	4	US-09-497-855A-44
C 9	33	6.0	399	4	US-09-621-976-8976
C 10	33	6.0	2135	4	US-08-933-71B-17
C 11	33	6.0	4612	3	US-08-960-048-2
C 12	33	6.0	4612	3	US-08-938-586-2
C 13	33	6.0	55827	4	US-09-813-133A-3
C 14	33	6.0	640681	4	US-09-790-988-1
C 15	33	6.0	1664976	4	US-08-916-421B-1
C 16	32.8	5.9	1751	4	US-09-453-702B-223
C 17	32.2	5.8	2169	4	US-09-434-408-3
C 18	32	5.8	2451	4	US-09-134-000C-2752
C 19	32	5.8	3336	4	US-09-717-364A-26
C 20	32	5.8	7291	3	US-08-913-943-3
C 21	32	5.8	26270	4	US-09-717-364A-1
C 22	31.8	5.8	1996	4	US-08-946-026-15
C 23	31.8	5.8	6124	4	US-08-213-419B-3
C 24	31.8	5.8	6773	4	US-09-166-350-27
C 25	31.6	5.7	1155	4	US-08-860-368B-19
C 26	31.4	5.7	696	4	US-09-543-681A-3120
C 27	31.2	5.6	3101	4	US-09-221-017B-87

C 28	31.2	5.6	3455	4	US-09-651-656-20	Sequence 20, App1
C 29	31.2	5.6	3455	4	US-09-650-855-20	Sequence 20, App1
C 30	31	5.6	876	3	US-08-531-927B-6	Sequence 6, App1
C 31	31	5.6	876	3	US-09-315-793-61	Sequence 61, App1
C 32	31	5.6	1512	3	US-08-945-994-4	Sequence 4, App1
C 33	31	5.6	1776	2	US-08-531-927B-1	Sequence 1, App1
C 34	31	5.6	1776	3	US-09-041-886-12	Sequence 12, App1
C 35	31	5.6	4176	4	US-09-722-139-1	Sequence 1, App1
C 36	31	5.6	4176	4	US-09-721-832-1	Sequence 1, App1
C 37	31	5.6	4176	4	US-09-721-832-1	Sequence 1, App1
C 38	31	5.6	6254	4	US-08-956-171B-404	Sequence 404, App
C 39	31	5.6	7015	3	US-09-177-249-6	Sequence 6, App1
C 40	31	5.6	14066	4	US-09-601-198-56	Sequence 56, App1
C 41	31	5.6	1664976	4	US-08-916-421B-1	Sequence 1, App1
C 42	31	5.6	1830121	4	US-09-557-884-1	Sequence 1, App1
C 43	31	5.6	1830121	4	US-09-643-990A-1	Sequence 1, App1
C 44	30.8	5.6	6617	4	US-09-976-594-268	Sequence 268, App
C 45	30.6	5.5	1047	4	US-09-543-681A-693	Sequence 693, App

ALIGNMENTS

RESULT 1
US-09-313-294A-1048/C
Sequence 1048, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Ialagudi, Raghunath V.
APPLICANT: Ito, Laura Y.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 1048
LENGTH: 277
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc. feature
OTHER INFORMATION: Incyte ID No. 6476212 700550303H1
US-09-313-294A-1048

Query Match 17.8%; Score 98.6; DB 4; Length 277;
Best Local Similarity 61.5%; Pred. No. 1.9e-19;
Matches 158; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY	167	ACTTTTCCCGAAGAGCTGTAAATAAGCTTCAAGTCGTTGGGCACTGTGTCTTCTT	226
DB	257	ATTTCCTCTGATGATGGCGCCAAAGACTTGAGATCTTGAGATCCGCTTCTTCAAT	198
QY	227	GATGAGATATCCCTTGTCTTCACTGCTTCACTCCAGCTGCAAGTATGATTTTCAAC	286
DB	197	CAGAGCGAGGCGCTTGATCGCGAAGCTTACTTGATCTTCAAGTATGATTTTCAAC	138
QY	287	CAAGGAGAGCTGTGGACATTAATGACCCCAATAAATCTTCTTCTTGTATGA	346
DB	137	CGAAGGAAAGCTGGGAATGTCTAATGAGCCCTTGTATCTTGTCTTCTTGTATGA	78
QY	347	TCCATTCCTTCTTCACTTTAAGTCACTCATATGATATACCAACAGTTCTTCTTCC	406
DB	77	ACCATTCACCGTAATCTCAAGCTTATGATTAATGATTAATCTTCTTCTTCTTCTG	18
QY	407	GAATATATCAATCAAT 423	
DB	17	GGACATATGACCTGAAT 1	

RESULT 2
US-08-232-463-14

APPLICANT: Huang, Tim
TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION
FILE REFERENCE: UMO1523
CURRENT APPLICATION NUMBER: US/09/497,855A
CURRENT FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/120,592
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: 60/118,760
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn version 3.0
SEQ ID NO: 37
LENGTH: 193303
TYPE: DNA
ORGANISM: Homo sapiens;
US-09-497-855A-37

Query Match 6.0%; Score 33.4; DB 4; Length 193303;
Best Local Similarity 45.1%; Pred. No. 22;
Matches 124; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Db 247 TCACCTGAGTCTACCTCCACCTGCAAGTCATCTAGTCCAAATGAGAACTGGGACA 306
141724 TTACGCTTAAGCTTAAGTGGCTTAATTAAGACATGAAACACATCATATCTTTCCAAA 141783
QY 307 TCATATGACCCCAATTAACCTTCTCTCTCTGATGATCCATCCCTTTGACTTT 366
141784 TGTTTGAACCTTAATATATATCTTCAATGTTTCACTTACTGAAATTTTGTTC 141843
QY 367 AAGGTCACTATATATATTAACCAACAGTTCTTCTGTCGACATATACCATGATCA 426
141844 ATGATATATCTTCAACCAAAATCTTCTTATCAATGATCAATGATCTTTCTTCCC 141903
QY 427 ACATTTTCTATATATCTGAACCTGACGAGTACCTGAAAAAATTAAGTGAAGAAC 486
141904 CTGCTGTAGCTTAAGATTAAGATTAAGATTTTCTTCAACAAAGTTGGGGGAGATC 141963
Db 487 TGATTTATGCGCAACGCAATTTGTTACATCTTTT 521
141964 TTATTTTAACATACTTGAATGATGTAATTTT 141998

RESULT 8
US-09-497-855A-44
Sequence 44, Application US/09497855A
Patent No. 6605432
GENERAL INFORMATION:
APPLICANT: Huang, Tim
TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION
FILE REFERENCE: UMO1523
CURRENT APPLICATION NUMBER: US/09/497,855A
CURRENT FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/120,592
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: 60/118,760
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn version 3.0
SEQ ID NO: 44
LENGTH: 193303
TYPE: DNA
ORGANISM: Homo sapiens;
US-09-497-855A-44

Query Match 6.0%; Score 33.4; DB 4; Length 193303;
Best Local Similarity 45.1%; Pred. No. 22;
Matches 124; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 247 TCACCTGAGTCTACCTCCACCTGCAAGTCATCTAGTCCAAATGAGAACTGGGACA 306
Db 141724 TTACGCTTAAGCTTAAGTGGCTTAATTAAGACATGAAACACATCATATCTTTCCAAA 141783
QY 307 TCATATGACCCCAATTAACCTTCTCTCTCTGATGATCCATCCCTTTGACTTTT 366

Db 141784 TGTTTGAACCTTAATATATATCTTCAATGTTTCACTACTTGAATTTTGTTC 141843
QY 367 AAGGTCACTATATATATTAACCAACAGTTCTTGTTCGACATATACCATGATCA 426
141844 ATGATATATCTTCAACCAAAATCTTCTTATCAATGATCAATGATCTTTCTTCCC 141903
QY 427 ACATTTTCTATATATCTGAACCTGACGAGTACCTGAAAAAATTTAGTGAAGAAC 486
141904 CTGCTGTAGCTTAAGATTAAGATTAAGATTTTCTTCAACAAAGTTGGGGGAGATC 141963
Db 487 TGATTTATGCGCAACGCAATTTGTTACATCTTTT 521
141964 TTATTTTAACATACTTGAATGATGTAATTTT 141998

RESULT 9
US-09-621-976-8976/c
Sequence 8976, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO: 8976
LENGTH: 399
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-8976

Query Match 6.0%; Score 33; DB 4; Length 399;
Best Local Similarity 15.3%; Pred. No. 2.7;
Matches 21; Conservative 66; Mismatches 48; Indels 0; Gaps 0;

QY 277 TCATGTTACCAATGAACTCTGGACATCTATATGACCCCAATTAATCTTCTCT 336
Db 209 WCTSRKMYKRYKRRKRRKCTSTKRYGISTYKWKAYTKRRRTWTYYSYM 150
QY 337 CCTGTATGATPCATTCCTTGAATTTAAGTCAACCATATGATTAACCAACAGT 396
Db 149 SMKTYMKTAIYWTKMKMTKTKWTCTWCKCTTYMAGTMMYRYRYRYAKRAKMS 90
QY 397 TTCTGTTCCGAATAT 413
Db 89 KRCWTSTCYCKMKMK 73

RESULT 10
US-08-933-711B-17
Sequence 17, Application US/08933711B
Patent No. 6514724
GENERAL INFORMATION:
APPLICANT: McMahon, Andrew P.
TITLE OF INVENTION: HEDGEHOG INTERACTING PROTEINS AND USES RELATED THERETO
FILE REFERENCE: HUV-024.01
CURRENT APPLICATION NUMBER: US/08/933,711B
CURRENT FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/026,155
PRIOR FILING DATE: 1996-09-20
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 17
LENGTH: 2135
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

